

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 23:24:18 ; Search time 46 Seconds
(without alignments)
1359,528 Million cell updates/sec

Title: US-10-057-832-2
Sequence: 1 MSREMDPVDIAEVKPLVEKG.....TPNSGAGNSAGPKSEVSC 394

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq.19Jun03.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2072	100.0	394	23	Human CAP43 protei
2	2072	100.0	394	24	Differentially exp
3	2067	99.8	394	23	Hypoxia-regulated
4	2067	99.8	394	24	Breast cancer asso
5	2067	99.8	394	24	Protein differenti
6	1250	60.3	375	23	Human NOV89B prote
7	1246	60.1	375	23	Human NOV89A prote
8	1231	59.4	363	22	Human reducing age
9	1229	59.3	363	23	Human NOV89c prote

10	1224	59.1	363	22	AAW94019	Human stomach can
11	1224	59.1	363	22	AAW94019	Human protein sequ
12	1158	55.9	359	23	ABU65168	Human NOV69d prote
13	1157	55.8	391	21	AAW23611	Human secreted pro
14	1137.5	54.9	339	22	AAW39870	Human polypeptide
15	1137.5	54.9	339	22	AAW39870	CDIF-17, Incyte I
16	1137.5	54.9	347	22	AAW41656	Human polypeptide
17	1102	53.2	251	21	AAW43959	Human cancer assoc
18	1049	50.6	371	23	AAU99505	Human Ndr2-related
19	1048	50.6	357	21	AAW41706	Human ORFX ORF1470
20	1048	50.6	357	23	ABW83808	Human P40. Homo s
21	1048	50.6	371	22	AAW79576	Human protein SEQ
22	1042	50.3	371	22	AAW78592	Human protein SEQ
23	1039	50.1	402	23	AAU99506	Human Ndr2-related
24	1039	50.1	575	23	AAU99507	Human Ndr2-related
25	936	45.2	337	23	ABW83809	Human P40-2. Homo
26	870	42.0	382	22	ABG10563	Novel human diagno
27	865	41.7	395	22	ABJ11598	Novel human secret
28	859	41.5	299	23	ABU65169	Human NOV89e prote
29	855	41.3	325	23	ABW89674	Human polypeptide
30	850	41.0	247	22	ABG16537	Novel human diagno
31	812.5	39.2	286	22	ABW95462	Novel human diagno
32	731	35.3	251	22	ABG02995	Novel human diagno
33	601.5	29.0	343	22	ABW67375	Drosophila melanog
34	601.5	29.0	425	22	ABW70926	Drosophila melanog
35	601.5	29.0	429	22	ABW67374	Drosophila melanog
36	579.5	28.0	202	23	ABW89673	Human polypeptide
37	511	24.7	730	22	ABG19705	Novel human diagno
38	511	24.7	745	22	ABG02993	Novel human diagno
39	498.5	24.1	209	22	ABG10562	Novel human diagno
40	492.5	23.8	209	22	AAU31599	Novel human secret
41	464	22.4	87	21	AAU00716	Human secreted pro
42	402	19.4	106	23	ABG77040	Prostate specific
43	402	19.4	113	21	AAU00715	Human secreted pro
44	362	17.5	355	22	ABW88816	Drosophila melanog
45	356.5	17.2	347	21	AAW29136	Arabidopsis thalia

ALIGNMENTS

RESULT 1	ABP52828	ABP52828 standard; Protein; 394 AA.
XX	ABP52828;	
AC	ABP52828;	
XX		
DT	31-OCT-2002 (first entry)	
XX		
DE	Human CAP43 protein SEQ ID NO:2.	
XX		
KW	Human, CAP43; cytosolic; antiproliferative; cell targeted therapy; cancer; melanoma; lymphoma; malignant fibrous histiocytoma; tumour; metastasis; dysproliferative change; dysplasia; angiogenesis; psoriasis; ischaemia; atherosclerosis; stroke; inflammatory.	
KW		
XX		
OS	Homo sapiens.	
XX		
PN	W0200258719-A1.	
XX		
PD	01-AUG-2002.	
XX		
PF	25-JAN-2002; 2002MO-US01891.	
XX		
PR	25-JAN-2001; 2001US-264268P.	
XX		
PA	(UTNY) UNIV NEW YORK STATE.	
XX		
PI	Costa M, Salnikow K, Yee H;	
XX		
DR	WPI; 2002-627386/67.	
XX		
DR	N-PSDB; ABQ74965.	
XX		

PT Identifying a diseased cell or tissue, e.g. associated with abnormal
PT CAP43 expression, useful for diagnosing or detecting cancer, comprises
PT detecting an elevated level of CAP43 nucleic acid or gene product cell
PT or tissue

XX Claim 3, Fig 1B, 11bp; English.

XX The present sequence represents human CAP43. The present invention
CC describes a method (M1) for identifying a diseased cell or tissue, where
CC the disease is associated with abnormal CAP43 expression. M1 comprises
CC detecting an elevated level of CAP43 nucleic acid or gene product in a
CC sample of cells or tissues from the individual. CAP43 has cytoskeletal
CC and antiproliferative activities and can be used in cell targeted therapy.
CC The method is useful for identifying diseased cell or tissues,
CC specifically a disease associated with abnormal expression of CAP43,
CC such as cancer, e.g. lung, colon, kidney, breast, or prostate cancer,
CC melanoma, lymphoma, or malignant fibrous histiocytoma. Compounds, which
CC specifically bind to CAP43, are useful for treating cancer. Nucleic acids
CC comprising CAP43 fragments are useful as oligonucleotide probes and
CC primers to detect and amplify other nucleic acid molecules encoding
CC CAP43 analogues and homologues. CAP43 nucleic acids and proteins may
CC also be used in detection methods, and in treatment of tumours,
CC dysproliferative changes (e.g. metaplasias and dysplasias), disorders
CC involving inappropriate cell or tissue growth augmented by angiogenesis,
CC or psoriasis. These may further be used to identify atherosclerosis in
CC cells, tissues or organs, ischaemia, stroke, or inflammatory tissues.

XX Sequence 394 AA;

Query Match 100.0%; Score 2072; DB 23; Length 394;
Best Local Similarity 100.0%; Pred. No. 3.3e-194;
Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSREMODVLAELVKPLVEKGETITGLDQEPVQEDITLHGSVHTLCGTPKGNRPVIL 60
DB 1 MSREMODVLAELVKPLVEKGETITGLDQEPVQEDITLHGSVHTLCGTPKGNRPVIL 60
QY 61 TYHDIGNMHKTCTYNPLFNVYEDMOEITQHFVAVCHVDAFGQDGAASFPAGYWPSPMDQLAE 120
DB 61 TYHDIGNMHKTCTYNPLFNVYEDMOEITQHFVAVCHVDAFGQDGAASFPAGYWPSPMDQLAE 120
QY 121 MLPGVLAQDFGLKSIIGMGTAGAYILTRFALNNPEWEGVLVINVNPCAEGMDMAASKI 180
DB 121 MLPGVLAQDFGLKSIIGMGTAGAYILTRFALNNPEWEGVLVINVNPCAEGMDMAASKI 180
QY 181 SGMTOALPDWVSHLFGKEEMQSNVEVHTYRQHIYNDMNPGLHLFINAYNSRDLEIE 240
DB 181 SGMTOALPDWVSHLFGKEEMQSNVEVHTYRQHIYNDMNPGLHLFINAYNSRDLEIE 240
QY 241 RMPGTHVTTLQCPALLVVDSSPAVDVAVECNSKLDPTKTTLLKMAADCGGLPQISQPAK 300
DB 241 RMPGTHVTTLQCPALLVVDSSPAVDVAVECNSKLDPTKTTLLKMAADCGGLPQISQPAK 300
QY 301 LAEAFKYFVQGMGYMPSASMTRLMRSRTASGSSVTSLDGTRSRSHTSEGTSSRSHTSEGT 360
DB 301 LAEAFKYFVQGMGYMPSASMTRLMRSRTASGSSVTSLDGTRSRSHTSEGTSSRSHTSEGT 360
QY 361 RSRSHTSSEGAHLDTIPNSGAAGNSAGPKSMEVSC 394
DB 361 RSRSHTSSEGAHLDTIPNSGAAGNSAGPKSMEVSC 394

RESULT 2

ABUS7628 standard; Protein; 394 AA.

ABUS7628;

09-APR-2003 (first entry)

Differentialially expressed breast cancer associated protein #15.

Breast cancer; differential gene expression; BC-cDNA;

KW breast cancer diagnosis; breast cancer monitoring;
KW breast cancer treatment; breast cancer staging.

OS Homo sapiens.

PN US2002156263-A1.

PD 24-OCT-2002.

PP 04-OCT-2001; 2001US-0974298.

PR 05-OCT-2000; 2000US-238331P.

PA (CHEN/) CHEN H.

PI Chen H;

DR WPI; 2003-182653/18.

XX New cDNAs, which are differentially expressed in (metastatic) breast
PT cancer useful for diagnosing or staging, breast cancer, or for
PT monitoring the treatment of breast cancer in an individual
XX Example; SEQ ID NO 81; 30pp; English.

CC The invention describes a combination of cDNAs (designated BC-cDNAs),
CC which are differentially expressed in breast cancer. The combination
CC includes 152 cDNA sequences, or their complements. The protein encoded
CC by any of these BC-cDNAs is useful for screening several molecules or
CC compounds to identify at least one ligand that specifically binds the
CC protein, producing or preparing polyclonal or monoclonal antibodies, or
CC purifying antibodies from a sample. The antibodies, which specifically
CC bind the protein differentially expressed in breast cancer is useful for
CC detecting the expression of a protein in a sample. The BC-cDNAs are
CC also useful for diagnosing, monitoring the treatment of, or staging,
CC breast cancer. This is the amino acid sequence of a differentially
CC expressed breast cancer associated protein.

CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from the US patent office at
CC seqdata.uspto.gov/sequence.html?DocID=20020156263.

XX Sequence 394 AA;

Query Match 100.0%; Score 2072; DB 24; Length 394;
Best Local Similarity 100.0%; Pred. No. 3.3e-194;
Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSREMODVLAELVKPLVEKGETITGLDQEPVQEDITLHGSVHTLCGTPKGNRPVIL 60
DB 1 MSREMODVLAELVKPLVEKGETITGLDQEPVQEDITLHGSVHTLCGTPKGNRPVIL 60
QY 61 TYHDIGNMHKTCTYNPLFNVYEDMOEITQHFVAVCHVDAFGQDGAASFPAGYWPSPMDQLAE 120
DB 61 TYHDIGNMHKTCTYNPLFNVYEDMOEITQHFVAVCHVDAFGQDGAASFPAGYWPSPMDQLAE 120
QY 121 MLPGVLAQDFGLKSIIGMGTAGAYILTRFALNNPEWEGVLVINVNPCAEGMDMAASKI 180
DB 121 MLPGVLAQDFGLKSIIGMGTAGAYILTRFALNNPEWEGVLVINVNPCAEGMDMAASKI 180
QY 181 SGMTOALPDWVSHLFGKEEMQSNVEVHTYRQHIYNDMNPGLHLFINAYNSRDLEIE 240
DB 181 SGMTOALPDWVSHLFGKEEMQSNVEVHTYRQHIYNDMNPGLHLFINAYNSRDLEIE 240
QY 241 RMPGTHVTTLQCPALLVVDSSPAVDVAVECNSKLDPTKTTLLKMAADCGGLPQISQPAK 300
DB 241 RMPGTHVTTLQCPALLVVDSSPAVDVAVECNSKLDPTKTTLLKMAADCGGLPQISQPAK 300
QY 301 LAEAFKYFVQGMGYMPSASMTRLMRSRTASGSSVTSLDGTRSRSHTSEGTSSRSHTSEGT 360
DB 301 LAEAFKYFVQGMGYMPSASMTRLMRSRTASGSSVTSLDGTRSRSHTSEGTSSRSHTSEGT 360
QY 361 RSRSHTSSEGAHLDTIPNSGAAGNSAGPKSMEVSC 394

Db 361 RSRSHTEGAHLDTTPNSGAAGNSAGPKSMEVSC 394

RESULT 3

ABP65134 ID ABP65134 standard; Protein; 394 AA.

AC ABP65134;

DT 12-NOV-2002 (first entry)

DE Hypoxia-regulated protein #8.

XX Cyclostatic; vasoregulator; tranquilizer; antiatherosclerotic; gene therapy;
 XX antiinflammatory; vulnary; gynecological; ophthalmological; vaccine;
 XX hypoxia; tumorigenesis; angiogenesis; apoptosis; cancer;
 XX ischaemic condition; reperfusion injury; retinopathy; neonatal stress;
 XX preclapemia; atherosclerosis; inflammatory condition; wound healing;
 XX inflammation; erythropoiesis; hair loss; human.

OS Homo sapiens.

PN WO200246465-A2.

PD 13-JUN-2002.

PE 10-DEC-2001; 2001WO-GB05458.

PR 08-DEC-2000; 2000GB-0030076.

PR 08-FEB-2001; 2001GB-0003156.

PR 25-OCT-2001; 2001GB-0025666.

XX (OXFO-) OXFORD BIOMEDICA UK LTD.

PI White J, Mundy CR, Ward NR, Krige D, Kingsman SM, Harris RA;
 PI Rayner MN;

DR WPI; 2002-627238/67.

PT Identifying a gene involved in disease for treating hypoxia-regulated
 PT conditions, comprises comparing the transcriptome/proteome of two cell
 PT types under different conditions and identifying a differentially
 PT regulated gene -

PS Claim 35; Page 335; 538pp; English.

XX The present invention relates to methods for identifying genes and
 CC proteins that are implicated in a specific disease or physiological
 CC condition. The method comprises comparing the transcriptome/proteome of a
 CC specialised cell type implicated in a disease or condition with that of a
 CC second specialised cell type, under two experimental conditions, and
 CC identifying a gene that is differentially regulated in the two
 CC specialised cell types under experimental conditions. ABV7873-ABV78116
 CC and ABP65061-ABP65257 were identified using the methods of the invention.
 CC The coding sequences and proteins are useful for treating a disease in a
 CC patient, for manufacture of a medicament for treating hypoxia-regulated
 CC conditions, and for regulating tumorigenesis, angiogenesis, apoptosis,
 CC biological response to hypoxia conditions, or hypoxic-associated
 CC pathology in a patient. The coding sequences and proteins are also useful
 CC for monitoring the therapeutic treatment of a disease or physiological
 CC condition, such as cancer, ischaemic conditions, reperfusion injury,
 CC retinopathy, neonatal stress, preclapemia, atherosclerosis, inflammatory
 CC conditions, wound healing, inflammation, erythropoiesis or hair loss.

XX Sequence 394 AA;

Query Match 99.8%; Score 2067; DB 23; Length 394;

Best Local Similarity 99.7%; Pred No. 1e-193;

Matches 393; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSREMODVLAELVKEVKGETITGLQEPDVQEDITLHGSVHTLCGTPKGNRPVIL 60

Db 1 MSREMODVLAELVKEVKGETITGLQEPDVQEDITLHGSVHTLCGTPKGNRPVIL 60
 QY TYHDIGNNHKTQVPLNFVEDMOBITQHPAVCHVDAPGODGAASPAGVYVPSMDOLAE 120
 Db 61 TYHDIGNNHKTQVPLNFVEDMOBITQHPAVCHVDAPGODGAASPAGVYVPSMDOLAE 120
 QY 121 MLPGLVQFGKLSIIGGTGAGAYILTRFALNBNEMVEGLVINVPCEBGMMDMAASKI 180
 Db 121 MLPGLVQFGKLSIIGGTGAGAYILTRFALNBNEMVEGLVINVPCEBGMMDMAASKI 180
 QY 181 SGWTOALPDVWVSHLFKEKEQSNVEVHTYRQHTVNDNPNGNHLFINAYNSRDLIE 240
 Db 181 SGWTOALPDVWVSHLFKEKEQSNVEVHTYRQHTVNDNPNGNHLFINAYNSRDLIE 240
 QY 241 RPPMGHTVTLQCPALVWGDSSPAVDVAVYECNSKLPPTKTTLLKMDCCGLPQISOPAK 300
 Db 241 RPPMGHTVTLQCPALVWGDSSPAVDVAVYECNSKLPPTKTTLLKMDCCGLPQISOPAK 300
 QY 301 LAEAFKTFVQGMGYPASMTRLMRSRTASGSSVTSLDGTRSRSHTEGTRSRSHTEGT 360
 Db 301 LAEAFKTFVQGMGYPASMTRLMRSRTASGSSVTSLDGTRSRSHTEGTRSRSHTEGT 360
 QY 361 RSRSHTEGAHLDTTPNSGAAGNSAGPKSMEVSC 394
 Db 361 RSRSHTEGAHLDTTPNSGAAGNSAGPKSMEVSC 394

RESULT 4

ID ABR47543 standard; Protein; 394 AA.

AC ABR47543;

DT 12-JUN-2003 (first entry)

DE Breast cancer associated protein sequence SEQ ID NO:324.

XX Human; breast cancer; cytostatic; gene therapy.

OS Homo sapiens.

PN WO2003004989-A2.

PD 16-JAN-2003.

PE 21-JUN-2002; 2002WO-US19669.

PR 21-JUN-2001; 2001US-299887P.

PR 18-JUL-2001; 2001US-301572P.

PR 25-SEP-2001; 2001US-326002P.

PR 05-MAR-2002; 2002US-362585P.

PR 14-MAY-2002; 2002US-380391P.

XX (MILL-) MILLENIUM PHARM INC.

PI Lillie J, Gannavarapu M, Glat K, Hoerh S, Kamatkar S, Mertens M;
 PI Monahan JB, Myer V, Wang Y, Xu Y, Zhao X, Meyers RE, Baat RC;
 PI Hototobagyi GN, Puzetal L, Meric F, Sahin A, Mills GB;

DR WPI; 2003-210381/20.

DR N-PSDB; ACC50241.

PT Breast cancer diagnosis or treatment by comparing the level of
 PT expression of a marker in a patient sample with that in the control
 PT non-breast cancer sample -

PS Claim 1; SEQ ID 324; 128pp; English.

XX The present invention describes a method for assessing whether a patient
 CC is afflicted with breast cancer. The method comprises comparing the level
 CC of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and
 CC ABR47386 to ABR47632) in a patient sample and the normal level of

expression of the marker in a control non-breast cancer sample, where a significant increase in the level of expression of the marker in the patient sample and the normal level is an indication that the patient is afflicted with breast cancer. The breast cancer associated sequences from the present invention have cytosolic activities and can be used in gene therapy. The method is useful for diagnosing and treating breast cancer.

N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPRO at ftp.wipro.int/pub/published_pcc_sequences.

Sequence 394 AA:

Query Match 99.8%; Score 2067; DB 24; Length 394;
Best Local Similarity 99.7%; Pred. No. 1e-193;
Matches 393; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSREMDDVDLAEVKPLVEKGETITGLQEPDVOEODIETLHGSVHTLCTGPKGNRPVIL 60
DB 1 MSREMDDVDLAEVKPLVEKGETITGLQEPDVOEODIETLHGSVHTLCTGPKGNRPVIL 60
QY 61 TYHDIGMNHKTCTNPLFNVEDMOEITTOHPAVCHVDAPGQDGAASFPAGMYPSMDQLAE 120
DB 61 TYHDIGMNHKTCTNPLFNVEDMOEITTOHPAVCHVDAPGQDGAASFPAGMYPSMDQLAE 120
QY 121 MLPGLVLOQFGLKSIIGMTGAGAYILTRFALNPNPEWEGVLINVPCEAGMDMAASKI 180
DB 121 MLPGLVLOQFGLKSIIGMTGAGAYILTRFALNPNPEWEGVLINVPCEAGMDMAASKI 180
QY 181 SGWTOALPDMVVSHLFGKEEMQSNVEVHTYRQHIIVNDMNPGLHLFINAYNSRRDLEIE 240
DB 181 SGWTOALPDMVVSHLFGKEEMQSNVEVHTYRQHIIVNDMNPGLHLFINAYNSRRDLEIE 240
QY 241 RMPGTHHTVTLQCPALLVVGDSPPAVDAVVECNKSLDPTKTTLLKMAOCGGLPOISQPAK 300
DB 241 RMPGTHHTVTLQCPALLVVGDSPPAVDAVVECNKSLDPTKTTLLKMAOCGGLPOISQPAK 300
QY 301 LAEAFKRYFVQGMGMPASMTRLMRSRTASGSSTSLDGTSSRSHTSEGTSSRSHTSEGT 360
DB 301 LAEAFKRYFVQGMGMPASMTRLMRSRTASGSSTSLDGTSSRSHTSEGTSSRSHTSEGT 360
QY 361 RSRSHTSAGAHLDITPPNSGAAGNSAGAPKSMESVSC 394
DB 361 RSRSHTSAGAHLDITPPNSGAAGNSAGAPKSMESVSC 394

RESULT 5
ABU07398 standard; Protein; 394 AA.

XX ABU07398;
XX AC ABU07398;
XX 28-JAN-2003 (first entry)

DE Protein differentially regulated in prostate cancer #1.
XX

KW Prostate cancer; gene expression; differential regulation;
KW molecular marker; drug target; cancer detection; cancer diagnosis;
KW cancer staging; cancer grading; cancer assessing; cancer monitoring.

XX Homo sapiens.

XX WO200281638-A2.

XX 17-OCT-2002.

XX 08-APR-2002; 2002WO-US10824.

XX 06-APR-2001; 2001US-281731P.

XX 06-APR-2001; 2001US-281732P.

XX (ORIG-) ORIGENE TECHNOLOGIES INC.

XX PA

PI Sun Z, Jay G;
XX WPI; 2003-058520/05.

PT Novel genes which are differentially regulated in prostate cancer,
PT useful for diagnosing prostate cancer in prostate tissue sample and
PT assessing therapeutic or preventive intervention in prostate cancer
PT patients -

PS Claim 1; Page 198-199; 416pp; English.

CC The invention describes genes (1) which are differentially regulated in
CC prostate cancer. (1) is useful for diagnosing a prostate cancer in a
CC sample comprising prostate tissue, which involves determining the number
CC of target genes which are differentially-regulated in the sample, where
CC the number is indicative of the probability that the sample comprises
CC prostate cancer. (1) is useful for assessing a therapeutic or preventive
CC intervention in a subject having a prostate cancer, which involves
CC determining the expression levels in a sample comprising prostate tissue
CC of target genes which are differentially-regulated in prostate cancer.
CC Preferably, the expression levels of at least 10 genes are determined.
CC (1) is also useful for identifying agents that modulate a biological
CC activity of a polypeptide differentially-regulated in prostate cancer
CC cells, which involves contacting a polypeptide differentially-regulated
CC in prostate cancer cells with a test agent under conditions effective for
CC determining whether the test agent modulates the biological activity.
CC (1) is useful as molecular markers, as drug targets, and for detecting,
CC diagnosing, staging, grading, assessing, monitoring, prognosticating,
CC preventing or treating, determining predisposition to diseases and
CC conditions especially relating to prostate cancer. (1) and its expression
CC products are used in the diagnostic test to assay for presence of cancer
CC e.g., in tissue sections, in biopsy sample, in total RNA, in lymph, in
CC blood etc. (1) is useful for assessing cancer e.g., to determine the type
CC of cancer, its stage of development, the nature of genetic defect, etc.
CC The polypeptide encoded by (1) can be used as target for therapy or drug
CC discovery. (1) can also be used for expressing the polypeptide and thus
CC for searching specific binding partners of the polypeptide. (1) is
CC useful in therapeutic applications to treat prostate cancer. The
CC identification of specific genes, and groups of genes, expressed in
CC pathways physiologically relevant to prostate cancer permits the
CC definition of functional and disease pathways and the delineation of
CC targets in these pathways which are useful in diagnostic, therapeutic,
CC and clinical applications. This is the amino acid sequence of a protein
CC differentially regulated in prostate cancer.

XX Sequence 394 AA:

Query Match 99.8%; Score 2067; DB 24; Length 394;
Best Local Similarity 99.7%; Pred. No. 1e-193;
Matches 393; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSREMDDVDLAEVKPLVEKGETITGLQEPDVOEODIETLHGSVHTLCTGPKGNRPVIL 60
DB 1 MSREMDDVDLAEVKPLVEKGETITGLQEPDVOEODIETLHGSVHTLCTGPKGNRPVIL 60
QY 61 TYHDIGMNHKTCTNPLFNVEDMOEITTOHPAVCHVDAPGQDGAASFPAGMYPSMDQLAE 120
DB 61 TYHDIGMNHKTCTNPLFNVEDMOEITTOHPAVCHVDAPGQDGAASFPAGMYPSMDQLAE 120
QY 121 MLPGLVLOQFGLKSIIGMTGAGAYILTRFALNPNPEWEGVLINVPCEAGMDMAASKI 180
DB 121 MLPGLVLOQFGLKSIIGMTGAGAYILTRFALNPNPEWEGVLINVPCEAGMDMAASKI 180
QY 181 SGWTOALPDMVVSHLFGKEEMQSNVEVHTYRQHIIVNDMNPGLHLFINAYNSRRDLEIE 240
DB 181 SGWTOALPDMVVSHLFGKEEMQSNVEVHTYRQHIIVNDMNPGLHLFINAYNSRRDLEIE 240
QY 241 RMPGTHHTVTLQCPALLVVGDSPPAVDAVVECNKSLDPTKTTLLKMAOCGGLPOISQPAK 300
DB 241 RMPGTHHTVTLQCPALLVVGDSPPAVDAVVECNKSLDPTKTTLLKMAOCGGLPOISQPAK 300
QY 301 LAEAFKRYFVQGMGMPASMTRLMRSRTASGSSTSLDGTSSRSHTSEGTSSRSHTSEGT 360

Db 301 LAARAFYPVQAGMGWMSASMTRLMMSRTASGSSVTSLLGTRSRSHTSEGTSRSHTSEGT 360

Qy 361 RSRSHTSEGAHLDTPTNSGAGNAGSPKSMVEVSC 394

Db 361 RSRSHTSEGAHLDTPTNSGAGNAGSPKSMVEVSC 394

RESULT 6
ABU65166
ID ABU65166 standard; Protein; 375 AA

PR	10-JUL-2001;	2001US-304354P.
PR	31-JUL-2001;	2001US-309198P.
PR	16-AUG-2001;	2001US-312903P.
PR	10-SEP-2001;	2001US-318462P.
PR	12-SEP-2001;	2001US-318770P.
PR	27-SEP-2001;	2001US-325430P.
PR	27-SEP-2001;	2001US-325681P.
PR	18-OCT-2001;	2001US-330380P.
PR	31-OCT-2001;	2001US-335301P.
PR	14-NOV-2001;	2001US-333172P.
PR	14-NOV-2001;	2001US-333271P.
PR	14-NOV-2001;	2001US-333272P.
PR	14-NOV-2001;	2001US-333184P.
PR	14-NOV-2001;	2001US-333272P.
PR	21-NOV-2001;	2001US-333094P.
PR	03-DEC-2001;	2001US-337426P.
PR	03-DEC-2001;	2001US-338092P.
PR	04-DEC-2001;	2001US-337185P.
PR	03-JAN-2002;	2002US-345705P.
PR	07-MAR-2002;	2002US-0092900.

```

Db      242 LQONNKSKTKLKSTLLVVDNSPAVEAVEGNSRLNPTITLLKMAOGGILPQVVGPK 301
Qy      301 LAEAFYFVGQGWMPSSAMTRLMRSRT-ASGSSVTSJLDCRTSRHTSEGTSRHTSEG 359
Db      302 LLEAFYFVGQGWMPSSAMTRLMRSRTSTSSJGSGSPPSRSVT-----SNQSDG 354
Qy      360 TRSRHTSEGAHLDTTPNSGAAGNSAGPKSMVEVC 394
Db      355 TQESCESP-----DVLDRH-----QTMVEVC 375

RESULT 7
ABU65165
ID      ABU65165 standard; Protein; 375 AA.
AC      ABU65165;
XX      20-MAY-2003 (first entry)
XX      Human NOV89a protein.
DE      Human NOV89a protein.
XX      NOVA; cytostatic; cardiant; antiarteriosclerotic; antiasthmatic; cancer;
KM      hypotensive; cardiomyopathy; bronchial asthma; gene therapy; vaccine;
KW      human.
XX      Homo sapiens.
XX      WO200272757-A2.
XX      19-SEP-2002.
XX      08-MAR-2002; 2002WO-US06908.
PF      08-MAR-2001; 2001US-274101P.
PR      08-MAR-2001; 2001US-274194P.
PR      08-MAR-2001; 2001US-274281P.
PR      09-MAR-2001; 2001US-274322P.
PR      09-MAR-2001; 2001US-274849P.
PR      12-MAR-2001; 2001US-275235P.
PR      13-MAR-2001; 2001US-275578P.
PR      13-MAR-2001; 2001US-275579P.
PR      13-MAR-2001; 2001US-275601P.
PR      14-MAR-2001; 2001US-276000P.
PR      16-MAR-2001; 2001US-276776P.
PR      19-MAR-2001; 2001US-276994P.
PR      20-MAR-2001; 2001US-277239P.
PR      20-MAR-2001; 2001US-277321P.
PR      21-MAR-2001; 2001US-277791P.
PR      22-MAR-2001; 2001US-277833P.
PR      23-MAR-2001; 2001US-278152P.
PR      26-MAR-2001; 2001US-278894P.
PR      27-MAR-2001; 2001US-278999P.
PR      27-MAR-2001; 2001US-279034P.
PR      28-MAR-2001; 2001US-279344P.
PR      30-MAR-2001; 2001US-277338P.
PR      30-MAR-2001; 2001US-279995P.
PR      30-MAR-2001; 2001US-280233P.
PR      02-APR-2001; 2001US-280822P.
PR      02-APR-2001; 2001US-280822P.
PR      04-APR-2001; 2001US-280900P.
PR      13-APR-2001; 2001US-281194P.
PR      13-APR-2001; 2001US-281675P.
PR      30-APR-2001; 2001US-287424P.
PR      02-MAY-2001; 2001US-288066P.
PR      03-MAY-2001; 2001US-288342P.
PR      15-MAY-2001; 2001US-288528P.
PR      16-MAY-2001; 2001US-291099P.
PR      16-MAY-2001; 2001US-291240P.
PR      30-MAY-2001; 2001US-294485P.
PR      31-MAY-2001; 2001US-294489P.

```

```

PR      31-MAY-2001; 2001US-294899P.
PR      18-JUN-2001; 2001US-299027P.
PR      19-JUN-2001; 2001US-299303P.
PR      19-JUN-2001; 2001US-299310P.
PR      10-JUL-2001; 2001US-304354P.
PR      31-JUL-2001; 2001US-309198P.
PR      16-AUG-2001; 2001US-312903P.
PR      10-SEP-2001; 2001US-318462P.
PR      12-SEP-2001; 2001US-318770P.
PR      27-SEP-2001; 2001US-325430P.
PR      27-SEP-2001; 2001US-325681P.
PR      18-OCT-2001; 2001US-330380P.
PR      31-OCT-2001; 2001US-335301P.
PR      14-NOV-2001; 2001US-332172P.
PR      14-NOV-2001; 2001US-332271P.
PR      14-NOV-2001; 2001US-332272P.
PR      14-NOV-2001; 2001US-333184P.
PR      14-NOV-2001; 2001US-333272P.
PR      21-NOV-2001; 2001US-332094P.
PR      03-DEC-2001; 2001US-337426P.
PR      03-DEC-2001; 2001US-338092P.
PR      04-DEC-2001; 2001US-337185P.
PR      03-JAN-2002; 2002US-345705P.
PR      07-MAR-2002; 2002US-0092900.
XX      (CURA-) CURAGEN CORP.
XX      Padigaru M, Spytek KA, Shenoy SG, Taupier RJ, Pena CEA, Li L;
PI      Zernuhen BD, Gusev V, Ji W, Gorman L, Miller CE, Kekuda R;
PI      Paturajan W, Gangoli E, Vernet CM, Guo X, Tchierne V;
PI      Fernandes ER, Casman SJ, Malyankar UM, Gerlach V, Liu Y;
PI      Anderson D, Spaderna SK, Catterton E, Burgess C, Leite M, Zhong H;
XX      Alabrook JP, Lepley DM, Rieger DK;
XX      WPI; 2002-723332/78.
DR      N-PSDB; ABX97132.
XX      NOVA polypeptides and polynucleotides, useful for preventing or
PT      treating a disorder associated with aberrant NOVA expression or
PT      activity e.g., cancer, hypertension, atherosclerosis, cardiomyopathy or
PT      bronchial asthma -
XX      Claim 1; Page 352; 1103pp; English.
XX      This invention describes novel human NOVA polypeptides which have
CC      cytostatic, cardiant, antiarteriosclerotic, antiasthmatic and
CC      hypotensive activity. Pharmaceutical compositions comprising the NOVA
CC      proteins or nucleic acid molecules or NOVA antibodies are useful for
CC      preventing or treating a disorder associated with aberrant NOVA
CC      expression or activity e.g. cancer, hypertension, atherosclerosis,
CC      cardiomyopathy or bronchial asthma. The products of the invention can
CC      be used for gene therapy or in a vaccine. ABU65041-ABU65218 represent
CC      the NOVA polypeptides encoded by ABX97008-ABX97185.
XX      Sequence 375 AA;
SQ      Query Match 60.1%; Score 1246; DB 23; Length 375;
      Best Local Similarity 61.5%; Pred. No.3,1e-113;
      Matches 243; Conservative 56; Mismatches 68; Indels 26; Gaps 6;
Qy      4 EMQDVDAEYKPLVERGETITGLQEFVQEDIEITLHGSVHTLCGTRKGNRPVILTYH 63
Db      3 ELQDVQLTEIKPLL-NDKNGTRNFQDFDCQEHDIETTHGVVHTYINGLRKGNRPVILTYH 61
Qy      64 DIGMNRKTCYNPLFNTYEDNOETQHVAVCHVDAPGQDGAASPAGYMPSPMDLAEMLP 123
Db      62 DIGLNKSCFNAPFNFEDMOEITQHVAVCHVDAPGQDGAASPPTQYPTMDLEMLP 121
Qy      124 GVLQDFGLKSLIGMGAGAGYILTRPALNPEWVEGLVILNVPACAGMWDMAASKISG 183
Db      122 PVLTHLSKSLIGIVGAGYILSRPALNPELVESGLVILNVPCKAGMDMAASKISGL 181
Qy      184 TQALPDMVVSHLFGEKEMOSNVEVHTYQROIIVDNKPGNHLIFINAYNSRDLIERPM 243

```

```

Db      182 TTNVVDIIIAHFGQELQANLDIQYRMHIAODINODNQLFLNSYNGRRDIEIRPI 241
Qy      244 PGTH---TVTLQCALIVGDSSPAVDVAVECNSKLDPTKTTLLKMAOCGLPQISOPAK 300
Db      242 LGQDNKSKTKLKSTLLVGDNSPAVEAVECNSRLNPINTTLKMAOCGLPQVOPGK 301
Qy      301 LAEAFKYFVQGMGYMPSASMTRLMRSRT-ASGSSVTSLDGTRSRSHTSSEGTGRSRSHTSSEG 359
Db      302 LTFKFKYFLOGMGYVPSASMTRLMRSKRTHTSSSLGSGESPFRSSTV-----SNQSDG 354
Qy      360 TRSRSHTSSEGAHLDTTPNSGAAGNSAGPKSMVEVSC 394
Db      355 TQESCESP-----DVLDRH-----QTMVEVSC 375

RESULT 8
AAG64392
ID      AAG64392 standard; Protein; 363 AA.
XX
AC      AAG64392;
XX
DT      04-OCT-2001 (first entry)
XX
DE      Human reducing agent and tunicamycin-responsive protein 40.
XX
KW      Human; reducing agent; tunicamycin-responsive protein 40; cytostatic;
KW      vitruclal; immunomodulatory; antiinflammatory; haemostatic;
KW      malignant tumour; haemopathy; HIV infection; immunological disease;
KW      inflammation.
XX
OS      Homo sapiens.
XX
PN      WO200155375-A1.
XX
PD      02-AUG-2001.
XX
PF      15-JAN-2001; 2001WO-CN00050.
XX
PR      26-JAN-2000; 2000CN-011517.
XX
PA      (BIOD-) BIODOR GENE TECHNOLOGY LTD SHANGHAI.
PI      Mao Y, Xie Y.
XX
XX      WPI; 2001-488795/53.
XX      N-PSDB; AAH73967.
XX
PT      Human reducing agent and tunicamycin-responsive protein 40 and encoded
PT      polynucleotide, applicable in diagnosis and treatment of malignant
PT      tumor, hemopathy, immunological diseases and various inflammations -
XX
PS      Claim 1; Page 36; 40pp; Chinese.
XX
CC      The present sequence is the protein sequence for human reducing agent and
CC      tunicamycin-responsive protein 40. The protein and its coding sequence
CC      are useful in the diagnosis and treatment of malignant tumour,
CC      haemopathy, HIV infection, immunological diseases and various
CC      inflammations.
XX
SQ      Sequence 363 AA;

Query Match      59.4%; Score 1231; DB 22; Length 363;
Best Local Similarity 60.5%; Pred. No. 8,6e-112;
Matches 239; Conservative 57; Mismatches 61; Indels 38; Gaps 6;
Qy      4 EMQDVDAEVRKPLVKEGETITSLQEPDVBODITFLGSHVYVTLGTPKGNRPVILTYH 63
Db      3 ELQDVQLBIRKLAND-----KENDITTHGVAVHTIRGLPKGNRPVILTYH 49
Qy      64 DIGMNHKTCTYNPLFNYEDMQEITQHPAVCHVDAPGQDGAASFPAGYMPSPMDLAEMLP 123
Db      50 DIGLNHKSCTFNAPFNFEDMQEITQHPAVCHVDAPGQDGAASFPAGYMPSPMDLAEMLP 109

```

```

Qy      124 GVLQGFELKSIIGWGTGAGAYILTRFALNPNPEWEGVLINVPNCAEGMDMAASKISGW 183
Db      110 SVLHLHLSKSIIGVGAGAYILSRFALNHPELVEGLVLINVPNCAAGWIDMAASKISGL 169
Qy      184 TQALPDVAVSHLFCKEEMQSNVEVHTYRQHIYNDMPGNILHFLINAYNSRRDIEIRPM 243
Db      170 TTNVVDIIIAHFGQELQANLDIQYRMHIAODINODNQLFLNSYNGRRDIEIRPI 229
Qy      244 PGTH---TVTLQCALIVGDSSPAVDVAVECNSKLDPTKTTLLKMAOCGLPQISOPAK 300
Db      230 LGQDNKSKTKLKSTLLVGDNSPAVEAVECNSRLNPINTTLKMAOCGLPQVOPGK 289
Qy      301 LAEAFKYFVQGMGYMPSASMTRLMRSRT-ASGSSVTSLDGTRSRSHTSSEGTGRSRSHTSSEG 359
Db      290 LTFKFKYFLOGMGYVPSASMTRLMRSKRTHTSSSLGSGESPFRSSTV-----SNQSDG 342
Qy      360 TRSRSHTSSEGAHLDTTPNSGAAGNSAGPKSMVEVSC 394
Db      343 TQESCESP-----DVLDRH-----QTMVEVSC 363

RESULT 9
ABU65167
ID      ABU65167 standard; Protein; 363 AA.
XX
AC      ABU65167;
XX
DT      20-MAY-2003 (first entry)
XX
DE      Human NOV89c protein.
XX
KW      NOVX; cytostatic; cardiac; antiarteriosclerotic; antiasthmatic; cancer;
KW      hypotensive; cardiomyopathy; bronchial asthma; gene therapy; vaccine;
KW      human.
XX
OS      Homo sapiens.
XX
PN      WO200272757-A2.
XX
PD      19-SEP-2002.
XX
PF      08-MAR-2002; 2002WO-US06908.
XX
PR      08-MAR-2001; 2001US-274101P.
PR      08-MAR-2001; 2001US-274194P.
PR      08-MAR-2001; 2001US-274281P.
PR      08-MAR-2001; 2001US-274322P.
PR      09-MAR-2001; 2001US-274849P.
PR      12-MAR-2001; 2001US-275235P.
PR      13-MAR-2001; 2001US-275578P.
PR      13-MAR-2001; 2001US-275579P.
PR      13-MAR-2001; 2001US-275601P.
PR      14-MAR-2001; 2001US-276000P.
PR      16-MAR-2001; 2001US-276776P.
PR      19-MAR-2001; 2001US-276994P.
PR      20-MAR-2001; 2001US-277239P.
PR      20-MAR-2001; 2001US-277321P.
PR      20-MAR-2001; 2001US-277327P.
PR      20-MAR-2001; 2001US-277327P.
PR      21-MAR-2001; 2001US-277791P.
PR      22-MAR-2001; 2001US-277833P.
PR      23-MAR-2001; 2001US-278152P.
PR      26-MAR-2001; 2001US-278894P.
PR      27-MAR-2001; 2001US-278999P.
PR      27-MAR-2001; 2001US-279036P.
PR      28-MAR-2001; 2001US-279344P.
PR      30-MAR-2001; 2001US-277338P.
PR      30-MAR-2001; 2001US-279995P.
PR      30-MAR-2001; 2001US-280233P.
PR      02-APR-2001; 2001US-280802P.
PR      02-APR-2001; 2001US-280822P.
PR      02-APR-2001; 2001US-280900P.
PR      04-APR-2001; 2001US-281194P.

```

PR 13-APR-2001; 2001US-283675P.
 PR 30-APR-2001; 2001US-287424P.
 PR 02-MAY-2001; 2001US-288066P.
 PR 03-MAY-2001; 2001US-288342P.
 PR 03-MAY-2001; 2001US-288528P.
 PR 15-MAY-2001; 2001US-291190P.
 PR 16-MAY-2001; 2001US-291099P.
 PR 16-MAY-2001; 2001US-291240P.
 PR 30-MAY-2001; 2001US-294485P.
 PR 31-MAY-2001; 2001US-294899P.
 PR 18-JUN-2001; 2001US-299027P.
 PR 19-JUN-2001; 2001US-299303P.
 PR 19-JUN-2001; 2001US-299310P.
 PR 10-JUL-2001; 2001US-304354P.
 PR 31-JUL-2001; 2001US-309198P.
 PR 16-AUG-2001; 2001US-312903P.
 PR 10-SEP-2001; 2001US-318462P.
 PR 12-SEP-2001; 2001US-318770P.
 PR 27-SEP-2001; 2001US-325430P.
 PR 27-SEP-2001; 2001US-325681P.
 PR 18-OCT-2001; 2001US-330380P.
 PR 31-OCT-2001; 2001US-335301P.
 PR 14-NOV-2001; 2001US-332172P.
 PR 14-NOV-2001; 2001US-332271P.
 PR 14-NOV-2001; 2001US-332272P.
 PR 14-NOV-2001; 2001US-333184P.
 PR 14-NOV-2001; 2001US-333272P.
 PR 21-NOV-2001; 2001US-332094P.
 PR 03-DEC-2001; 2001US-337426P.
 PR 03-DEC-2001; 2001US-338092P.
 PR 04-DEC-2001; 2001US-337185P.
 PR 03-JAN-2002; 2002US-345705P.
 PR 07-MAR-2002; 2002US-0092900.
 PA (CURA-) CURAGEN CORP.
 XX
 PI Padigar M, Spytek KA, Shenoy SG, Taupier RJ, Pena CRA, Li L;
 PI Zernusen BD, Gusev V, Ji W, Gorman L, Miller CE, Kehuda R;
 PI Paturajan M, Gangoli E, Vermet CAM, Guo X, Tchenev V;
 PI Fernandes ER, Casman SJ, Malyankar UM, Gerlach V, Liu Y;
 PI Anderson D, Spaderan SK, Catterton E, Burgess C, Lette M, Zhong H;
 PI Alsobrook JP, Lepley DM, Rieger DK;
 XX
 DR MPI: 2002-723332/78.
 DR N-PSDB; ABX97134.
 XX
 PT NOVA polypeptides and polynucleotides, useful for preventing or
 PT treating a disorder associated with aberrant NOVA expression or
 PT activity e.g., cancer, hypertension, atherosclerosis, cardiomyopathy or
 PT bronchial asthma -
 XX
 PS Claim 1; Page 353; 1103pp; English.
 XX
 CC This invention describes novel human NOVA polypeptides which have
 CC cytosolic, cardiant, antiarteriosclerotic, antiasthmatic and
 CC hypotensive activity. Pharmaceutical compositions comprising the NOVA
 CC proteins or nucleic acid molecules or NOVA antibodies are useful for
 CC preventing or treating a disorder associated with aberrant NOVA
 CC expression or activity e.g. cancer, hypertension, atherosclerosis,
 CC cardiomyopathy or bronchial asthma. The products of the invention can
 CC be used for gene therapy or in a vaccine. ABU5041-ABU6518 represent
 CC the NOVA polypeptides encoded by ABX97008-ABX97185.
 CC
 XX
 SQ Sequence 363 AA;
 Query Match 59.3%; Score 1229; DB 23; Length 363;
 Best Local Similarity 60.5%; Pred. No. 1,3e-111;
 Matches 239; Conservative 57; Mismatches 61; Indels 38; Gaps 6;
 QY 4 EMOVDLAELVRLVKEGKITGLQEPDVGQDITLHSGVHTLCTGPKGNRPVILTYH 63
 DB 3 ELQDVQLREIKRLAND-----KEHDIETHTGVVHTIRGLPKGNRPVILTYH 49

QY 64 DIGMNHKTCVNPFLNVEYEDMOEITQHPAVCHVDAPGQODGASPPAGYPMQDLAEMLP 123
 DB 50 DIGMNHKTCVNPFLNVEYEDMOEITQHPAVCHVDAPGQODGASPPAGYPMQDLAEMLP 109
 QY 124 GVLQOFGIKSITIGMGAGAYILTRPALNPNEMVEGLVLINVPCEAGMDMAASKISGW 183
 DB 110 PVLTHLSLKSIIGIGVAGAYILSRFALNHPVEGLVLINVPCEAGMDMAASKISGL 169
 QY 184 TQALPDWVSHLREKEMQSNVEVHTYRQIYVNDMPGULHFINAYNSRDLERPM 243
 DB 170 TTNVVDIILAHFQOEELQANLUDIQTFRMHQDIDQDMLOFLNSYNRDLERPI 229
 QY 244 PGTH---TVTLQCPALIVGDPSPAPVAVECNSKLDPTTTLTKMADCGGLPOLISQPAK 300
 DB 230 LGQNDKSKTKLCTSLVVDGNSPAVAVECNSRLNPIITLLKMDCCGLPQVQDPK 289
 QY 301 LAEAFKYFVGMGVPSASMTRLRSRT-ASGSSVTLSDGTRSRSHTSBGRSRSHTSSEG 359
 DB 290 LTFEAFKYFLOGMGVPSASMTRLRSRTHTSTSSLSGSESPFERSVT-----SNQSDG 342
 QY 360 TRSRSHTSBGAHLDITPNSGAGNSAGPKSMVSC 394
 DB 343 TQESCESP-----DVLDRH-----QTMVEVSC 363
 RESULT 10
 AAM94019
 ID AAM94019 standard; Protein; 363 AA.
 AC AAM94019;
 XX
 DT 13-NOV-2001 (first entry)
 XX
 DE Human stomach cancer expressed polypeptide SEQ ID NO 108.
 XX
 KW Human; stomach cancer; marker; screening; micro-metastasis;
 KW peritoneal dissemination.
 XX
 OS Homo sapiens.
 XX
 PN WO200109317-A1.
 XX
 PD 08-FEB-2001.
 XX
 XX 28-JUL-2000; 2000MO-JP05063.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 18-OCT-1999; 99US-0159590.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 17-FEB-2000; 2000US-0183322.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T, Aburatani H;
 PI Kodama T, Midorikawa Y;
 XX
 DR MPI: 2001-570287/64.
 DR N-PSDB; AAI93896.
 XX
 PT New Stomach cancer-associated genes, useful as markers in blood tests
 PT for screening for the early stages of the disease -
 XX
 PS Claim 1; Page 189-190; 242pp; Japanese.
 XX
 CC The invention relates to stomach cancer-expressed genes
 CC (AAI93842-AAI93917) and the encoded proteins (AAM93967-AAM94039). The
 CC genes can be used as markers in blood tests for screening for the early
 CC stages of the disease. The proteins and peptides can be used as targets

CC for screening for compounds to treat the disease. They can also be used
CC for predicting micro-metastases. The gene can predict peritoneal
CC dissemination.

XX Sequence 363 AA:

Query Match 59.1%; Score 1224; DB 22; Length 363;
Best Local Similarity 60.5%; Pred. No. 4.2e-111;
Matches 239; Conservative 56; Mismatches 62; Indels 38; Gaps 6;

QY 4 EMQDDVLAEVKPLVEKGETITGLQDFVDQODIETLHGSVYVTLGTPKGRPVILTYH 63
DB 3 ELQDVQLTEIKPLND-----KEDIETTHGVVHTIRGLPKGRPVILTYH 49
QY 64 DIGMNHKTCYNPLFNYEDMOEITQHFVAVCHVDAPQODGAASPPAGYVPSMDOLAEMP 123
DB 50 DIGLNHKSCSNAPFNEDMOEITQHFVAVCHVDAPQODGAASPPAGYVPTMDELAEMP 109
QY 124 GYLQDFGLKSIIGMGTGAGAYILTRFALNPEVBEGLVILVNPFCAGMMDMAASKISGM 183
DB 110 PVLTHLSKSIIGIGVAGAYILSRFALNHPVLEGLVILVNDPCAKGMDMAASKISGL 169
QY 184 TQALPDVMSHLFGKEEMQSNVEVHTTRQHIYVNDMPGNLHLFTINANSRDLIERPM 243
DB 170 TTNVVDITIAHFGQEBELQANDLITQYRMHIAODINODNLQFLINSYNGRRDLIERPI 229
QY 244 PGTH---TVTLQCPALLVVDGSSPAVDVAVECNKLDPTKTLTKMADCGGLPQVQPGK 300
DB 230 LGQDNKSKTKLCSTLLVVDGNSPAVEAVECNSRLNPITTLTKMADCGGLPQVQPGK 289
QY 301 LAEAFKYFVQGMGMPASMTRLMRSRT-ASGSSVTSIDGTRSRSHTEGTRSRSHTEG 359
DB 290 LTEAFKYFLQMGYIPASMTRLASRTHSTSSLSGSESPFSRSVT-----SNQSDG 342
QY 360 TRSRSHTEGAHLDTTPNSGAAGSAGPKSMVEVC 394
DB 343 TOESCESP-----DVLDRH-----QTMEVSC 363

RESULT 11

AB94494
ID AAB94494 standard; Protein; 363 AA.

XX AAB94494;

AC 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:15186.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Oca T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Iehi S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -

XX Claim 8; SEQ ID 15186; 2537bp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification, where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC of an oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 363 AA:

Query Match 59.1%; Score 1224; DB 22; Length 363;
Best Local Similarity 60.5%; Pred. No. 4.2e-111;
Matches 239; Conservative 56; Mismatches 62; Indels 38; Gaps 6;

QY 4 EMQDDVLAEVKPLVEKGETITGLQDFVDQODIETLHGSVYVTLGTPKGRPVILTYH 63
DB 3 ELQDVQLTEIKPLND-----KEDIETTHGVVHTIRGLPKGRPVILTYH 49
QY 64 DIGMNHKTCYNPLFNYEDMOEITQHFVAVCHVDAPQODGAASPPAGYVPSMDOLAEMP 123
DB 50 DIGLNHKSCSNAPFNEDMOEITQHFVAVCHVDAPQODGAASPPAGYVPTMDELAEMP 109
QY 124 GYLQDFGLKSIIGMGTGAGAYILTRFALNPEVBEGLVILVNPFCAGMMDMAASKISGM 183
DB 110 PVLTHLSKSIIGIGVAGAYILSRFALNHPVLEGLVILVNDPCAKGMDMAASKISGL 169
QY 184 TQALPDVMSHLFGKEEMQSNVEVHTTRQHIYVNDMPGNLHLFTINANSRDLIERPM 243
DB 170 TTNVVDITIAHFGQEBELQANDLITQYRMHIAODINODNLQFLINSYNGRRDLIERPI 229
QY 244 PGTH---TVTLQCPALLVVDGSSPAVDVAVECNKLDPTKTLTKMADCGGLPQVQPGK 300
DB 230 LGQDNKSKTKLCSTLLVVDGNSPAVEAVECNSRLNPITTLTKMADCGGLPQVQPGK 289
QY 301 LAEAFKYFVQGMGMPASMTRLMRSRT-ASGSSVTSIDGTRSRSHTEGTRSRSHTEG 359
DB 290 LTEAFKYFLQMGYIPASMTRLASRTHSTSSLSGSESPFSRSVT-----SNQSDG 342
QY 360 TRSRSHTEGAHLDTTPNSGAAGSAGPKSMVEVC 394
DB 343 TOESCESP-----DVLDRH-----QTMEVSC 363

RESULT 12

ABU65168
ID ABU65168 standard; Protein; 359 AA.

XX ABU65168;

AC 20-MAY-2003 (first entry)

DE Human NOV99d protein.

04-DEC-2001; 2001US-337185P.
03-JAN-2002; 2002US-345705P.
07-MAR-2002; 2002US-0092900.

(CURA-) CURAGEN CORP.

Padigaru M, Spytek KA, Shenoy SG, Taupier RJ, Pena CEA, Li L; Zernusen BD, Guev V, Ji W, Gorman L, Miller CE, Kerkuda R; Patusturjan M, Gangoli E, Vernet CM, Guo X, Tchener V; Pi Ferrandez ER, Casman SJ, Malyanar UM, Gerlach V, Liu Y; Pi Anderson D, Spaderna SK, Caterton E, Burgess C, Leite M, Zhong H; Alsbrook JP, Lepley DM, Rieger DK;

WPI; 2002-723332/78.
N-PSDB; ABX97135.

NOVX polypeptides and polynucleotides, useful for preventing or treating a disorder associated with aberrant NOVX expression or activity e.g., cancer, hypertension, atherosclerosis, cardiomyopathy or bronchial asthma -

Claim 1, Page 354; 1103pp; English.

This invention describes novel human NOVX polypeptides which have cyostatic, cardiant, antiarteriosclerotic, antiasthmatic and hypotensive activity. Pharmaceutical compositions comprising the NOVX proteins or nucleic acid molecules or NOVX antibodies are useful for preventing or treating a disorder associated with aberrant NOVX expression or activity e.g. cancer, hypertension, atherosclerosis, cardiomyopathy or bronchial asthma. The products of the invention can be used for gene therapy or in a vaccine. AB05041-AB055218 represent the NOVX polypeptides encoded by ABX97008-ABX97185.

Sequence 359 AA;

Query Match 55.9%; Score 1158; DB 23; Length 359;
Best Local Similarity 58.2%; Pred. No. 1.2e-104;
Matches 230; Conservative 55; Mismatches 68; Indels 42; Gaps 7;

4 EMODVDAEYVLEKKGTTGLOEFPYVODIETLHGSVHVLGCPKGNRPVILTYH 63
3 ELQDVLEIKFPL-NDKNGTRNFQDFPCOEHDLETHGVVHTIRGLPKGNRPVILTYH 61
64 DIGNHNKTCVNPFLFNYEDMOETITQFVAVCHYDAGQODGASFPAGYVYPSMDQLAEMLP 123
62 DIGNRRKSCFNAFPNFMEDMOETITQFVAVCHYDAGQODGASFPFTGYQYPTMDLAEMLP 121
124 GYLOQFGIKSTIGMGTGAGAYITRFALNPNEMVEGLVLINVPQABGMDMAASKISGW 183
122 PVLTHLSKSTIGGVGAGAYILSRFALNHPLEVEGLVLINVDCAKGMIDMAASKISGL 181
184 TQALPDWVSHLPEKEMQSNVEVHVRHRIVNDMNGNLFLINAVNSRDLERPM 243
182 ITTNVVDIILAHFQOEELQANVLDIYRNMHTADINODNQLPLNSNGRDLIEPI 241
244 PGTH---TVTLQCCALLVWGDSBPADVAVVECSKLDPTKTTLLKMACDGLPQISOPAK 300
242 LGQNDKSKTLKCGTLLVVGDSNAYAV-----MADCGGLPVVQPGK 285
301 LAEAFKYFVQGMGYPSASMTRLMRST-ASGSSVTSLDGTRSRHTSEGTRSRHTSEG 359
286 FTBAFKYFLQMGYTPASMTRLMRSTHSTSSLSGSGESPFSRVT-----SNQSDG 336
360 TRSRHTSEGAHLDITNSGAAAGSAPKSNVEVC 394
339 TQESCESP-----DVLDRL-----QTMVEVC 359

RESULT 13
AAB23611
ID AAB23611 standard; Protein; 391 AA.
XX AAB23611;
XX

```

XX 12-JAN-2001 (first entry)
XX
XX Human secreted protein SEQ ID NO: 22.
XX
XX Human; secreted protein; cytokine; cell proliferation;
XX nutritional supplement; immune modulation; autoimmune disorder;
XX haematopoiesis regulation; tissue growth; haemostasis; inflammation.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Peptide 10..22
XX Protein /label= signal_peptide
XX /label= mature_protein
XX
XX WO200049134-A1.
XX
XX 24-AUG-2000.
XX
XX 18-FEB-2000; 2000WO-US04340.
XX
XX 19-FEB-1999; 99US-0120680.
XX 23-APR-1999; 99US-0298733.
XX 17-AUG-1999; 99US-0149639.
XX 23-SEP-1999; 99US-0155686.
XX 01-OCT-1999; 99US-0157247.
XX 29-NOV-1999; 99US-0167822.
XX 29-NOV-1999; 99US-0167823.
XX 15-FEB-2000; 2000US-0298733.
XX
XX (ALPH-) ALPHAGENE INC.
XX
XX Valenzuela D, Yuan O, Hoffman H, Hall J, Rapiejko P,
XX
XX WPI; 2000-549267/50.
XX N-PSDB; AAA93111.
XX
XX New secreted proteins and polynucleotides encoding them, which are
XX derived from Homo sapiens, useful for therapy, diagnosis, and research,
XX as well as nutritional sources or supplements -
XX
XX Claim 31; Page 255-257; 309pp; English.
XX
XX The present sequence is the sequence of a human secreted protein. Its
XX cDNA was isolated from an adult pancreas cDNA library. The proteins
XX and coding sequences of the invention can be used in the isolation of
XX similar genes and proteins, in the elucidation of their function in vivo,
XX and to treat a number of conditions. It is possible that they may have
XX uses as nutritional supplements, as cytokine or cell proliferation
XX factors, in immune modulation, where they may be used to treat immune and
XX autoimmune diseases, as haematopoiesis regulators (treating myeloid or
XX lymphoid cell deficiencies), in the promotion of tissue growth, they may
XX have chemokine or chemotactic activity, haemostatic or thrombolytic
XX activity, or anti-inflammatory activity.
XX
XX Sequence 391 AA:
SQ

```

```

Query Match 55.8%; Score 1157; DB 21; Length 391;
Best Local Similarity 56.3%; Pred. No. 1.7e-104;
Matches 224; Conservative 58; Mismatches 78; Indels 38; Gaps 4;

```

```

QY 5 MODVLAELVPLVEKGTITG-----LLOEPVQEDDITLHGSHVTLTGPKGNR 56
DB 24 LQELRPFEPKPL-EGODATBLESSDAFLADTWMKHDITPYGLHVVIRGSPKGNR 82
QY 57 PILTYHDIGMNHKTCVNPENYEDMOETTHFAVCHVDAPQOGAASFPAGWYPSMD 116
DB 83 PAILTYHDIGMNHKTCVNPENYEDMOETTHFAVCHVDAPQOGAASFPAGWYPSMD 142
QY 117 QLAEMLPGLVQOFGKSLIGMGTGAGAYLITRPAIINPEMVGVLINVPACBGMMDWA 176

```

```

DB 143 QLAEMLPGLVQOFGKSLIGMGTGAGAYLITRPAIINPEMVGVLINVPACBGMMDWA 202
QY 177 ASKISGWTQALPDMVSHLPGKEMOSNVVHTYRQIYVDNMGNHLPTINYSRRD 236
DB 203 ATKLSGLSTLPDLVSHLSQBELVNNTELVOSYRQOIGVAVDANLQLEPMNYSRRD 262
QY 237 LEIERPMPGTHATVTLACFPALIVGDSPPAVDAVECNKSLDPTKTYLLKMAIDCGGLPDIS 296
DB 263 LDINRPGVPAKTLRCPVNLVVGDNAPAEQGVNENSKLDPPTTTFLKMAIDSGGLPQVT 322
QY 297 QPAKLBAFKYFVQGMGMPASMTLRKRTASGSSVTSIDGTRSRSHTSRGTRSRSH 356
DB 323 QPGKLTAFKYLQGMGMPASMTLRKRTASGSSVTSIDGTRSRSHTSRGTRSRSH 376
QY 357 SEGRSRSHTSRGTRSRSHTSRGTRSRSHTSRGTRSRSHTSRGTRSRSHTSRGTRSRSH 394
DB 377 SEGIGQVNHNT-----NEVSC 391

RESULT 14
AAM39870
ID AAM39870 standard; Protein; 339 AA.
XX
XX AAM39870;
XX
XX 22-OCT-2001 (first entry)
XX
XX Human polypeptide SEQ ID NO 3015.
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.
XX
XX Homo sapiens.
XX
XX WO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YF, Liu C, Aundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
XX Zhao QH, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI; 2001-442253/47.
XX N-PSDB; AAI59026.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Example 4; SEQ ID NO 3015; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and
XX the encoded polypeptides (AAM38642-AA42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and

```



```

Db      128 PDLVEGLVLVNIIDPNGKGMIDMAATKLSGLTSTLPDVTLSHLPSQEBELVNNTELVSQYRQ 187
Qy      214 HIYNDMPGNLHLFINAYNSRRDLIERPMPGTHVTILOCPALVVGDSPPAVDAVBCN 273
Db      188 QIGNVVNOANLQLFMMNMTNSRDLIDINRPGTVPNAKTILACFVMLVVGDNAPADGVBCN 247
Qy      274 SKLDPTKTTLLMADCGGLPQISQPAKLAFAFKYFVQGWGYPASMTLMSRTASGSS 333
Db      248 SKLDPTTTTFLKMDSGGLPQVTPGKLTFAFKYFLOGWGYPASMTLMSRTASLTS 307
Qy      334 VTSLECTSRSHSTSECTSRSHSTSECTSRSHSTSECAHLDTPNAGAGNSAGPKSMEVS 393
Db      308 ASSVDGSRPQACT-----HSESSEGLQGVNHT-----MEVS 338
Qy      394 C 394
Db      339 C 339

```

Search completed: January 29, 2004, 23:27:54
 Job time : 48 secs


```
QY 181 SGMTOALPDMVYSHLFGKEEMOSNVEVHTYRQHIYNDMPGNLHLFINAYSRRDLIE 240
DB 181 SGMTOALPDMVYSHLFGKEEMOSNVEVHTYRQHIYNDMPGNLHLFINAYSRRDLIE 240
QY 241 RMPGTHVTTLQCPALLVVGDSPPAVDAVECNKSLDPTKTLTKMADCGGLPQISQPAK 300
DB 241 RMPGTHVTTLQCPALLVVGDSPPAVDAVECNKSLDPTKTLTKMADCGGLPQISQPAK 300
QY 301 LAEAFKTFVQGMGMPASMTRLMRSRTASGSSTSLDGTSSRSHTSRSHTSRSGT 360
DB 301 LAEAFKTFVQGMGMPASMTRLMRSRTASGSSTSLDGTSSRSHTSRSHTSRSGT 360
QY 361 RSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSGT 394
DB 361 RSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSGT 394
```

```
RESULT 2
US-09-919-039-197
; Sequence 197, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaset, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; PRIOR FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 197
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 1969721CD1
US-09-919-039-197
```

```
Query Match 100.0%; Score 2072; DB 11; Length 394;
Best Local Similarity 100.0%; Pred. No. 5.2e-199;
Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSREMODVDLAEVKPLVEKGETITGLQEPDVQODIETLHGSVHTLCGTPKGNRPVIL 60
DB 1 MSREMODVDLAEVKPLVEKGETITGLQEPDVQODIETLHGSVHTLCGTPKGNRPVIL 60
QY 61 TYHDIGNMHTCYNPLFNVEDMOEITQHFVACHVDAPGQODGAASFPAGYMPSMDQLAE 120
DB 61 TYHDIGNMHTCYNPLFNVEDMOEITQHFVACHVDAPGQODGAASFPAGYMPSMDQLAE 120
QY 121 MLPGLVLOQFGKLSITIGMGTGAGAYILTRPALNPNMEVGLVILINVPACAGMDMAASKI 180
DB 121 MLPGLVLOQFGKLSITIGMGTGAGAYILTRPALNPNMEVGLVILINVPACAGMDMAASKI 180
QY 181 SGMTOALPDMVYSHLFGKEEMOSNVEVHTYRQHIYNDMPGNLHLFINAYSRRDLIE 240
DB 181 SGMTOALPDMVYSHLFGKEEMOSNVEVHTYRQHIYNDMPGNLHLFINAYSRRDLIE 240
QY 241 RMPGTHVTTLQCPALLVVGDSPPAVDAVECNKSLDPTKTLTKMADCGGLPQISQPAK 300
DB 241 RMPGTHVTTLQCPALLVVGDSPPAVDAVECNKSLDPTKTLTKMADCGGLPQISQPAK 300
QY 301 LAEAFKTFVQGMGMPASMTRLMRSRTASGSSTSLDGTSSRSHTSRSHTSRSGT 360
DB 301 LAEAFKTFVQGMGMPASMTRLMRSRTASGSSTSLDGTSSRSHTSRSHTSRSGT 360
QY 361 RSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSGT 394
DB 361 RSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSGT 394
```

```
RESULT 3
US-10-172-069-33
; Sequence 33, Application US/10172069
; Publication No. US20030167480A1
; GENERAL INFORMATION:
; APPLICANT: Stuart, Susan G.
; APPLICANT: Au-Young, Janice
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Shah, Puryi
; APPLICANT: Yue, Henry
; TITLE OF INVENTION: NDR2-RELATED PROTEINS
; FILE REFERENCE: PC-0038 CIP
; CURRENT APPLICATION NUMBER: US/10/172,069
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US/09/812,484
; PRIOR FILING DATE: 2001-03-19
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 33
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Genbank ID No. US20030167480A1 g1596167
US-10-172-069-33
```

```
Query Match 100.0%; Score 2072; DB 12; Length 394;
Best Local Similarity 100.0%; Pred. No. 5.2e-199;
Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSREMODVDLAEVKPLVEKGETITGLQEPDVQODIETLHGSVHTLCGTPKGNRPVIL 60
DB 1 MSREMODVDLAEVKPLVEKGETITGLQEPDVQODIETLHGSVHTLCGTPKGNRPVIL 60
QY 61 TYHDIGNMHTCYNPLFNVEDMOEITQHFVACHVDAPGQODGAASFPAGYMPSMDQLAE 120
DB 61 TYHDIGNMHTCYNPLFNVEDMOEITQHFVACHVDAPGQODGAASFPAGYMPSMDQLAE 120
QY 121 MLPGLVLOQFGKLSITIGMGTGAGAYILTRPALNPNMEVGLVILINVPACAGMDMAASKI 180
DB 121 MLPGLVLOQFGKLSITIGMGTGAGAYILTRPALNPNMEVGLVILINVPACAGMDMAASKI 180
QY 181 SGMTOALPDMVYSHLFGKEEMOSNVEVHTYRQHIYNDMPGNLHLFINAYSRRDLIE 240
DB 181 SGMTOALPDMVYSHLFGKEEMOSNVEVHTYRQHIYNDMPGNLHLFINAYSRRDLIE 240
QY 241 RMPGTHVTTLQCPALLVVGDSPPAVDAVECNKSLDPTKTLTKMADCGGLPQISQPAK 300
DB 241 RMPGTHVTTLQCPALLVVGDSPPAVDAVECNKSLDPTKTLTKMADCGGLPQISQPAK 300
QY 301 LAEAFKTFVQGMGMPASMTRLMRSRTASGSSTSLDGTSSRSHTSRSHTSRSGT 360
DB 301 LAEAFKTFVQGMGMPASMTRLMRSRTASGSSTSLDGTSSRSHTSRSHTSRSGT 360
QY 361 RSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSGT 394
DB 361 RSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSHTSRSGT 394
```

```
RESULT 4
US-10-057-832-2
; Sequence 2, Application US/10057832
; Publication No. US20020182621A1
; GENERAL INFORMATION:
; APPLICANT: Costa, Max
; APPLICANT: Salnikow, Konstantin
; APPLICANT: Yee, Herman
; TITLE OF INVENTION: METHODS AND COMPOSITIONS USING CA93 PROTEINS AND NUCLEIC ACIDS TO
; FILE REFERENCE: 5986/11147051
; CURRENT APPLICATION NUMBER: US/10/057,832
```


ORGANISM: Mus musculus
 FEATURE:
 OTHER INFORMATION: Cytoplasmic protein Ndr1
 US-10-205-194-156

Query Match 94.6%; Score 1960; DB 12; Length 394;
 Best Local Similarity 93.7%; Pred. No. 9e-188;
 Matches 369; Conservative 14; Mismatches 11; Indels 0; Gaps 0;

QY 1 MSREMODVDLAELVKEKGETITGLQEPVQEDIDITLHGSVHTLCGTPKGNRPVIL 60
 DB 1 MSREIHVDLAELVKEKGETITGLQEPVQEDIDITLHGSVHTLCGTPKGNRPVIL 60
 QY 61 TYHDIGMNHKTCYNPLFNYEDMOEITQHFAVCHVADAPGQDGAASFPAGYMPSMDQLAE 120
 DB 61 TYHDIGMNHKTCYNPLFNYEDMOEITQHFAVCHVADAPGQDGAASFPAGYMPSMDQLAE 120
 QY 121 MLPGVLOOFGLSKSIIGMGTGAGAYILTRPALNPNEMVEGLVILNPNCAEGMMDMAASKI 180
 DB 121 MLPGVLOOFGLSKSIIGMGTGAGAYILTRPALNPNEMVEGLVILNPNCAEGMMDMAASKI 180
 QY 181 SGMTOALPDMVSHLFGKEEMQSNVEVHTYRQHIIVNDMPGNLHLFINAYNSRRDLEIE 240
 DB 181 SGMTOALPDMVSHLFGKEEIHNNVEVHTYRQHIIVNDMPGNLHLFINAYNSRRDLEIE 240
 QY 241 RPMGTHVTTLQCPALLVVDSSPAVDVVECNKLDPTKTTLLKMACCGGLPQISQPAK 300
 DB 241 RPMGTHVTTLQCPALLVVDGNSPAVDVVECNKLDPTKTTLLKMACCGGLPQISQPAK 300
 QY 301 LAEAFKTYVQGMGMPASMTRLMRSRTASGSVTSLDGTRSRSHTSSEGT 360
 DB 301 LAEAFKTYVQGMGMPASMTRLMRSRTASGSVTSLEGTSRSHTSSEGT 360
 QY 361 RSRSHTSSEGAHLDTTPNSGAAGNSAGPKSMEVSC 394
 DB 361 RSRSHTSSEGAHLDTTPNSGATGNAGPKSMEVSC 394

RESULT 7
 US-10-108-260A-3803
 ; Sequence 3803, Application US/10108260A
 ; Publication No. US20040005560A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HELIX RESEARCH INSTITUTE
 ; TITLE OF INVENTION: NO. US20040005560A1e1 full length cDNA
 ; FILE REFERENCE: HI-A0106
 ; CURRENT APPLICATION NUMBER: US/10/108, 260A
 ; CURRENT FILING DATE: 2002-03-27
 ; NUMBER OF SEQ ID NOS: 5458
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 3803
 ; LENGTH: 356
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-108-260A-3803

Query Match 87.7%; Score 1817; DB 12; Length 356;
 Best Local Similarity 90.1%; Pred. No. 1.7e-173;
 Matches 355; Conservative 0; Mismatches 1; Indels 38; Gaps 1;

QY 1 MSREMODVDLAELVKEKGETITGLQEPVQEDIDITLHGSVHTLCGTPKGNRPVIL 60
 DB 1 MSREMODVDLAELVKEKGETITGLQEPVQEDIDITLHGSVHTLCGTPKGNRPVIL 60
 QY 61 TYHDIGMNHKTCYNPLFNYEDMOEITQHFAVCHVADAPGQDGAASFPAGYMPSMDQLAE 120
 DB 61 TYHDIGMNHKTCYNPLFNYEDMOEITQHFAVCHVADAPGQDGAASFPAGYMPSMDQLAE 120
 QY 121 MLPGVLOOFGLSKSIIGMGTGAGAYILTRPALNPNEMVEGLVILNPNCAEGMMDMAASKI 180
 DB 121 MLPGVLOOFGLSKSIIGMGTGAGAYILTRPALNPNEMVEGLVILNPNCAEGMMDMAASKI 180
 QY 181 SGMTOALPDMVSHLFGKEEMQSNVEVHTYRQHIIVNDMPGNLHLFINAYNSRRDLEIE 240

DB 143 SGMTOALPDMVSHLFGKEEMQSNVEVHTYRQHIIVNDMPGNLHLFINAYNSRRDLEIE 202
 QY 241 RPMGTHVTTLQCPALLVVDSSPAVDVVECNKLDPTKTTLLKMACCGGLPQISQPAK 300
 DB 203 RPMGTHVTTLQCPALLVVDSSPAVDVVECNKLDPTKTTLLKMACCGGLPQISQPAK 262
 QY 301 LAEAFKTYVQGMGMPASMTRLMRSRTASGSVTSLDGTRSRSHTSSEGT 360
 DB 263 LAEAFKTYVQGMGMPASMTRLMRSRTASGSVTSLDGTRSRSHTSSEGT 322
 QY 361 RSRSHTSSEGAHLDTTPNSGAAGNSAGPKSMEVSC 394
 DB 323 RSRSHTSSEGAHLDTTPNSGAAGNSAGPKSMEVSC 356

RESULT 8
 US-10-108-260A-3274
 ; Sequence 3274, Application US/10108260A
 ; Publication No. US20040005560A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HELIX RESEARCH INSTITUTE
 ; TITLE OF INVENTION: NO. US20040005560A1e1 full length cDNA
 ; FILE REFERENCE: HI-A0106
 ; CURRENT APPLICATION NUMBER: US/10/108, 260A
 ; CURRENT FILING DATE: 2002-03-27
 ; NUMBER OF SEQ ID NOS: 5458
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 3274
 ; LENGTH: 323
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-108-260A-3274

Query Match 79.5%; Score 1646.5; DB 12; Length 323;
 Best Local Similarity 82.0%; Pred. No. 1.9e-156;
 Matches 323; Conservative 0; Mismatches 0; Indels 71; Gaps 1;

QY 1 MSREMODVDLAELVKEKGETITGLQEPVQEDIDITLHGSVHTLCGTPKGNRPVIL 60
 DB 1 MSREMODVDLAELVKEKGETITGLQEPVQEDIDITLHGSVHTLCGTPKGNRPVIL 60
 QY 61 TYHDIGMNHKTCYNPLFNYEDMOEITQHFAVCHVADAPGQDGAASFPAGYMPSMDQLAE 120
 DB 61 TYHDIGMNHKTCYNPLFNYEDMOEITQHFAVCHVADAPGQDGAASFPAGYMPSMDQLAE 120
 QY 121 MLPGVLOOFGLSKSIIGMGTGAGAYILTRPALNPNEMVEGLVILNPNCAEGMMDMAASKI 180
 DB 121 MLPGVLOOFGLSKSIIGMGTGAGAYILTRPALNPNEMVEGLVILNPNCAEGMMDMAASKI 180
 QY 181 SGMTOALPDMVSHLFGKEEMQSNVEVHTYRQHIIVNDMPGNLHLFINAYNSRRDLEIE 240
 DB 181 SGMTOALPDMVSHLFGKEEMQSNVEVHTYRQHIIVNDMPGNLHLFINAYNSRRDLEIE 240
 QY 241 RPMGTHVTTLQCPALLVVDSSPAVDVVECNKLDPTKTTLLKMACCGGLPQISQPAK 300
 DB 241 RPMGTHVTTLQCPALLVVDSSPAVDVVECNKLDPTKTTLLKMACCGGLPQISQPAK 300
 QY 301 LAEAFKTYVQGMGMPASMTRLMRSRTASGSVTSLDGTRSRSHTSSEGT 360
 DB 301 LAEAFKTYVQGMGMPASMTRLMRSRTASGSVTSLDGTRSRSHTSSEGT 360
 QY 361 RSRSHTSSEGAHLDTTPNSGAAGNSAGPKSMEVSC 394
 DB 290 RSRSHTSSEGAHLDTTPNSGAAGNSAGPKSMEVSC 323

RESULT 9
 US-10-094-749-1842
 ; Sequence 1842, Application US/10094749
 ; Publication No. US20030219741A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ISOGAL, TAKAO

RESULT 10
 US-09-925-301-1404
 ; Sequence 1404, Application US/09925301
 ; Patent No. US2005005300A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rosen et al.
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

```

RESULT 11
US-10-172-069-1
1 Sequence 1, Application US/10172069
2 Publication No. US20030167480A1
3 GENERAL INFORMATION:
4 APPLICANT: Stuart, Susan G.
5 APPLICANT: Au-Young, Janice
6 APPLICANT: Hillman, Jennifer L.
7 APPLICANT: Spah, Purvi
8 APPLICANT: Yue, Henry
9 TITLE OF INVENTION: NDR-RELATED PROTEINS
10 FILE REFERENCE: PC-0038 CIP
11 CURRENT APPLICATION NUMBER: US/10/172,069
12 CURRENT FILING DATE: 2002-06-13
13 PRIOR APPLICATION NUMBER: US/09/812,484
14 PRIOR FILING DATE: 2001-03-19
15 NUMBER OF SEQ. ID NOS: 33
16 SOFTWARE: PERL Program
17 SEQ ID NO 1
18 LENGTH: 371
19 TYPE: PRT
20 ORGANISM: Homo sapiens
21 FEATURE:
22 NAME/KEY: misc_feature
23 OTHER INFORMATION: Incyte ID No. US20030167480A1 2227688CDP1
24 US-10-172-069-1

```

Query Match	50.6%	Score 1049;	DB 12;	Length 371;
Best Local Similarity	52.1%	Pred. No. 2.1e-96;		
Matches 208; Conservative	64;	Mismatches 89;	Indels 38;	Gaps 6;

QY 4 EMOVDLAELVPLV-----EKGETITGLQEPFVQEDODIETLHGSVHTLGGTPKGN 55
 Db 3 ELQEVQITREKRLPGQTPBEAKELARILLD-QGQTHSVETPYGVTFTVGTPEKPK 61
 QY 56 RPVILTYHDIQGNHKTCPNPLFNVEDMOETIOHPAVCHVADPQODGAASFPAGVYPSM 115
 Db 62 RPAILTYHDVGLNYSKCFQPLFQEPDMOEIIQNFVAVHVDAPGMEGAPVFPGLGYPSL 121
 QY 116 DQLEMLPGVLOQFGLKSIIGMGTGAGAYILTRFALNPNVEVGLVILNPNCAEGMMDW 175
 Db 122 DQLDMLPCVLOQLNFTIIGVGAGAYILARVALNPDVBEGLVILNIDPNKGMMDW 181
 QY 176 AASKISGWTQALPDMVNVSHLFGKEEMQSNVEVHTYRQHIYVNDMPGNLH.FINAYNSRR 235
 Db 182 AAHLTGLTSSIPDMILGHLFSQBELSGNSELIQKRYNIITHAPLNDIEIYNSYNRR 241
 QY 236 DLEIRPMPGTHVTTLQCPALLVVGDSPPAVDAVECNKLDPTKTTLLKMADEGGLPOI 295
 Db 242 DLNFER---GGDITLRCPVMLVVGDDQPHEDAVVECNKLDPTQTSFLKMADEGGLPOI 297
 QY 296 SOPAKLAELAFKYFVQGMGYMPSASMTRLMRSTRASGSSVTSLDGTRSRSH 355
 Db 298 TOPGLTEAFKYFVQGMGYMSSCMTRLSRSTRASGSSVTSLDGTRSRSH 346
 QY 356 TSEGRSRSHTEGALHLDITPNSGAGNSAGPKSMVESC 394
 Db 347 -----RTLSQSSSESG-----TLSSGPGRH-----TMEVSC 371

RESULT 12
 US-10-172-069-2
 ; Sequence 2, Application US/10172069
 ; Publication No. US20030167480A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Stuart, Susan G.
 ; APPLICANT: Au-Young, Janice
 ; APPLICANT: Hillman, Jennifer L.
 ; APPLICANT: Shah, Purvi
 ; APPLICANT: Yue, Henry
 ; TITLE OF INVENTION: NDR2-RELATED PROTEINS
 ; FILE REFERENCE: PC-0038 CIP
 ; CURRENT APPLICATION NUMBER: US/10/172,069
 ; PRIOR FILING DATE: 2002-06-13
 ; PRIOR APPLICATION NUMBER: US/09/812,484
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 2
 ; LENGTH: 402
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; OTHER INFORMATION: incycle ID No. US20030167480A1 3507515CD1
 ; NAME/KEY: unsure
 ; LOCATION: 328, 374, 379
 ; OTHER INFORMATION: unknown or other
 ; US-10-172-069-2

Query Match 50.1%; Score 1039; DB 12; Length 402;
 Best Local Similarity 55.1%; Pred. No. 2,4e-95;
 Matches 201; Conservative 59; Mismatches 85; Indels 20; Gaps 5;
 QY 4 EMOVDLAELVPLV-----EKGETITGLQEPFVQEDODIETLHGSVHTLGGTPKGNRPVILTYH 63
 Db 3 ELQEVQITREKRLPGQTPBEAKELARILLD-QGQTHSVETPYGVTFTVGTPEKPKPALITYH 55
 QY 64 DIGNHHKTCPNPLFNVEDMOETIOHPAVCHVADPQODGAASFPAGVYPSMQLAMLP 123
 Db 56 DVLNYSKCFQPLFQEPDMOEIIQNFVAVHVDAPGMEGAPVFPGLGYPSLQDLMDIP 115

QY 124 GVLQOFLAKSIIGMGTGAGAYILTRFALNPNVEVGLVILNPNCAEGMMDWAAKISGW 183
 Db 116 CVIQLNFTIIGVGAGAYILARVALNPDVBEGLVILNIDPNKGMMDWAAHLKLTGL 175
 QY 184 TQALPDMVNVSHLFGKEEMQSNVEVHTYRQHIYVNDMPGNLH.FINAYNSRRDIETIRPM 243
 Db 176 TSSIPDMILGHLFSQBELSGNSELIQKRYNIITHAPLNDIEIYNSYNRRDLNFER-- 233
 QY 244 PGHTVTTLQCPALLVVGDSPPAVDAVECNKLDPTKTTLLKMADEGGLPOIQAPKLA 303
 Db 234 --GGDITLRCPVMLVVGDDQPHEDAVVECNKLDPTQTSFLKMADEGGLPOITOPGLT 291
 QY 304 AFKYFVQGMGYMPSASMTRLMRSTRASGSSVTSLDGTRSRSH-SEGRS-----RS 354
 Db 292 AFKYFVQGMGYMSSCMTRLSRSTRASGSSVTSLDGTRSRSH 351
 QY 355 HTSBG 359
 Db 352 HTWGG 356

RESULT 13
 US-10-172-069-32
 ; Sequence 32, Application US/10172069
 ; Publication No. US20030167480A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Stuart, Susan G.
 ; APPLICANT: Au-Young, Janice
 ; APPLICANT: Hillman, Jennifer L.
 ; APPLICANT: Shah, Purvi
 ; APPLICANT: Yue, Henry
 ; TITLE OF INVENTION: NDR2-RELATED PROTEINS
 ; FILE REFERENCE: PC-0038 CIP
 ; CURRENT APPLICATION NUMBER: US/10/172,069
 ; PRIOR FILING DATE: 2002-06-13
 ; PRIOR APPLICATION NUMBER: US/09/812,484
 ; NUMBER OF SEQ ID NOS: 33
 ; SOFTWARE: PERL Program
 ; SEQ ID NO 32
 ; LENGTH: 371
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; OTHER INFORMATION: Genbank ID No. US20030167480A1 96141566
 ; US-10-172-069-32

Query Match 50.1%; Score 1038; DB 12; Length 371;
 Best Local Similarity 52.1%; Pred. No. 2,7e-95;
 Matches 208; Conservative 64; Mismatches 89; Indels 38; Gaps 7;
 QY 4 EMOVDLAELVPLV-----EKGETITGLQEPFVQEDODIETLHGSVHTLGGTPKGN 55
 Db 3 ELQEVQITREKRLPGQTPBEAKELARILLD-QGQTHSVETPYGVTFTVGTPEKPK 61
 QY 56 RPVILTYHDIQGNHKTCPNPLFNVEDMOETIOHPAVCHVADPQODGAASFPAGVYPSM 115
 Db 62 RPAILTYHDVGLNYSKCFQPLFQEPDMOEIIQNFVAVHVDAPGMEGAPVFPGLGYPSL 121
 QY 116 DQLEMLPGVLOQFGLKSIIGMGTGAGAYILTRFALNPNVEVGLVILNPNCAEGMMDW 175
 Db 122 DQLDMLPCVLOQLNFTIIGVGAGAYILARVALNPDVBEGLVILNIDPNKGMMDW 181
 QY 176 AASKISGWTQALPDMVNVSHLFGKEEMQSNVEVHTYRQHIYVNDMPGNLH.FINAYNSRR 235
 Db 182 AAHLTGLTSSIPDMILGHLFSQBELSGNSELIQKRYNIITHAPLNDIEIYNSYNRR 241
 QY 236 DLEIRPMPGTHVTTLQCPALLVVGDSPPAVDAVECNKLDPTKTTLLKMADEGGLPOI 295
 Db 242 DLNFER---GGDITLRCPVMLVVGDDQPHEDAVVECNKLDPTQTSFLKMADEGGLPOI 297
 QY 296 SOPAKLAELAFKYFVQGMGYMPSASMTRLMRSTRASGSSVTSLDGTRSRSH 355

[illegible]

```

RESULT 14
US-10-104-047-2056
: Sequence 2056, Application US/10104047
: Publication NO. US20030236392A1
: GENERAL INFORMATION:
: APPLICANT: HELIX RESEARCH INSTITUTE
: TITLE OF INVENTION: NO. US20030236392A1 full length cdna
: FILE REFERENCE: H1-A0105
: CURRENT APPLICATION NUMBER: US/10/104, 047
: CURRENT FILING DATE: 2002-03-25
: PRIOR APPLICATION NUMBER:
: PRIOR FILING DATE:
: NUMBER OF SEQ ID NOS: 4096
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 2056
: LENGTH: 311
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-10-104-047-2056

```

Query Match	43.1%	Score	892.5	DB	12	Length	311
Best Local Similarity	55.0%	Pred	No. 8.3e-81				
Matches	176	Conservative	48	Mismatches	67	Indels	29
						Gaps	4

[illegible]

RESULT 15
US-10-264-237--2050
Sequence 2050, Application US/10264237
Publication No. US20040009491A1
GENERAL INFORMATION:
APPLICANT: Birse et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PA131P1
CURRENT APPLICATION NUMBER: US/10/264,237
CURRENT FILING DATE: 2002-10-04
PRIOR APPLICATION NUMBER: PCT/US01/16450
PRIOR FILING DATE: 2001-05-18
PRIOR APPLICATION NUMBER: US 60/205,515
PRIOR FILING DATE: 2000-05-19
NUMBER OF SEQ ID NOS: 2876
SOFTWARE: PatentIn Ver. 3.1

```

? SEQ ID NO 2050
? LENGTH: 325
? TYPE: PRT
? ORGANISM: Homo sapiens
? FEATURE:
? NAME/KEY: MISC FEATURE
? LOCATION: (285)
? OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
?-S-10-264-237-2030

```

Query Match	41.3%	Score 855;	DB 12;	Length 325;
Best Local Similarity	51.9%;	Pred. No. 5.1e-77;		
Matches 163;	Conservative 59;	Mismatches 74;	Indels 18;	Gaps 5

```

QY      4 EMOVDVLAIVKRLPVEVGGATITLLOEFPVQRODLETLGSHVHVLGGPKSNRVLITVH 63
Db      3 ELQEOVITEKPLLT- PGQTPEA----- AKTHSVETTPGASVTFPVYGGPKPKRRALITVH 55

QY      64 DIGMNHKTGYNPLFNVEDMOEITOHFAVCHVADAPQOOGAASPAGVYVPEMDLAELP 123
Db      56 DVLANIKSGFQELPQEPEDMOEIIQNFYKVYVHVDAPMEGAVVFPFGIQYPSLDQADMI 115

QY      124 GVLQOFGKLSIIIGMGTGAGAYLITRFPALNBNMEVGLVINVNCAGGMDMAASKISGW 183
Db      116 CVLQYINFPSTIIIGVGAGAYILARYALNHDPVEGLVLINIDNAGGMDMAAHKLTGL 175

QY      184 TQALPMMVYSHFGKEBMOSNVYVHTROHLYVDMVNGVHLPLTVNSRRDLIERPM 243
Db      176 TSSIPENIIIGHLFQSOBELSGNSELIQKTRNIITTHAPLNDIELVWNSYNNRRDLNFR-- 233

QY      244 PGTHVTVLCOPALLVVGSSPAVDAVEBNSKLDPTKTLTKMADCGGLPOLISQPAKLA 303
Db      234 --GSDITFLCIPVMLVVGDDAPHEBDAVVEBNSKLDPTQTSFLKMADSGGQPOLTX-ARQAD 290

QY      304 AFKTYVQSGMGMP 317
Db      291 -----RGLQVLP 298

```

Search completed: January 29, 2004, 23:35:00
Job time : 41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 23:26:14 / Search time 21 Seconds
(without alignments)
1804.308 Million cell updates/sec

Title: US-10-057-832-2

Perfect score: 2072
Sequence: 1 MSRMQVDVLAEVKFLVEKG.....TPNSGAAGNSAGPKSEVSC 394

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	356.5	17.2	347	2	AB4579
2	351.5	17.0	352	2	T13993
3	347	16.7	325	2	T27688
4	312	15.1	361	2	T48538
5	149	7.2	335	2	D83263
6	140	6.8	315	2	H83276
7	138.5	6.7	256	1	BVECH
8	138.5	6.7	275	2	AB1655
9	135.5	6.5	393	2	B87548
10	134.5	6.5	256	2	P91160
11	134.5	6.5	256	2	B86006
12	128.5	6.2	275	2	AB1284
13	127.5	6.2	255	2	C82692
14	125	6.0	256	2	AF0016
15	123	5.9	258	2	AE0997
16	119	5.7	287	2	AF2333
17	114.5	5.5	268	2	G82042
18	112	5.4	267	2	B87535
19	111.5	5.4	267	2	C89873
20	111.5	5.4	625	2	T41603
21	110.5	5.3	269	2	T36625
22	107.5	5.2	1361	2	S29998
23	107	5.2	340	2	AD3040
24	107	5.2	340	2	H98245
25	107	5.2	1217	2	T25894
26	106.5	5.1	1788	2	T31095
27	103.5	5.0	1810	1	A32230
28	103.5	5.0	273	2	E70010
29	103	5.0	273	2	E70010

30	103	5.0	275	2	AC2293	hypothetical prote
31	102.5	4.9	286	2	AG3220	hydrolyase [impor
32	102	4.9	302	2	D70538	probable hpoB prot
33	102	4.9	683	2	AB2704	1,4-beta-cellubios
34	101.5	4.9	265	2	AB3115	beta-ketoadipate e
35	101.5	4.9	265	2	C98172	beta-ketoadipate e
36	101.5	4.9	781	2	H84501	En/Spm-like transp
37	100.5	4.9	265	2	F83586	probable hydrolyase
38	100.5	4.9	267	2	P90516	esterase/lipase 1
39	100.5	4.9	275	2	AF1951	hypothetical prote
40	99.5	4.8	313	2	E83543	probable hydrolyase
41	99.5	4.8	659	2	C96730	probable ABC trans
42	99	4.8	523	1	S17949	glutamate dehydrog
43	98.5	4.8	270	2	T38218	probable abhydrola
44	98.5	4.8	273	2	E90516	esterase/lipase 1
45	98.5	4.8	1362	2	A37474	surface glycoprote

ALIGNMENTS

```
RESULT 1
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: AB4420; MUID:20083487; PMID:10617197
A:Accession: AB4579
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-347 <STO>
A:Cross-references: GB:AE002093; NID:G4191791; PION:AAID0160.1; GSPDB:GN00139
A:Gene: At2g19620
A:Map position: 2

Query Match      17.2%; Score 356.5; DB 2; Length 347;
Best Local Similarity 30.3%; Pred. No. 1.4e-21;
Matches 90; Conservative 59; Mismatches 125; Indels 23; Gaps 7;

QY 33 QEDDIETLHGSVHTLCGTPKGNRPVILTVHDIQNMKTCYPLFNYEDMQEITQH-FAV 91
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 20 KENHVTKCHGSVSVVVVGDE--KPLITVPDALVNMSCFQGLFCLPBAVSLLNHPCI 77

QY 92 CHVDAPGQDGAASFPAQGYRPPSMDLAEMLPVYLQDFGLKSIIGMGAGAYLITPAL 151
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 78 YHISPPGHEFGAIVCSNDSPSVEDLADQILELVNFFSLEAVCMGKITGAYLISFAI 137

QY 152 NNPMVEGLVLINVPACAEGMMDMAASKI-----SGWTQALPDVVVSHLFGKEBQSN 204
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 138 KHERVVLGLILISPLCAQPSMSEFFYKVVSNLLYYMGSLKIDILQYFSEAGSS 197

QY 205 V---EVVHTYRQHIIVNDNPNGLHLFINAYSNRDLIERPMPGHTTATVTLQCPALLVG 260
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 198 EEPEDVAVHCRR-LIGERHSSLMRFLBAVNRHDLT-----DLKSLKCTLLIFVG 249

QY 261 DSGPAVDVAVECNSKLDPTTTLTKAADCGGLQIISPAFLAEAFKFFVVGNG- YMP 316
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 250 DQSPFHSRTELMTVYALDRKYSALVEVACSGMVTBEQPHAMLIPEEFFMGFGLYRP 306

RESULT 2
T13993
hypothetical protein sf21 - common sunflower
C/Species: Helianthus annuus (common sunflower)
C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C/Accession: T13993
```


C/Accession: D83263
 R/Stover, C.K.; Plam, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A/Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho
 A/Reference number: A82950; MUID:20437337; PMID:10984043
 A/Accession: D83263
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-335 <STO>
 A/Cross-references: GB:AE004730; GB:AE004091; NID:99949154; PIDN:AA06441.1; GSPDB:GN001
 A/Experimental source: strain PA01
 C/Genetics:
 A/Gene: PA3953

Query Match 7.2%; Score 149; DB 2; Length 335;
 Best Local Similarity 25.1%; Pred. No. 0.00017;
 Matches 63; Conservative 43; Mismatches 85; Indels 60; Gaps 14;

Qy 104 ASPAGWYPMDDLAEMLPGLVLOQFGKLSITGCTGAGAYILTRPALNPEWEGVLI 163
 Db 107 SSKPAHYY-SFOQLAANTHALERGVASVIGSHSGMLATRYALLVPROVERLV- 164
 Qy 164 NVNFCAGGMDMAASKISGWTQALPDVVVSHLFGKEMQSNVEVHTYRQ-HIVDMNP 221
 Db 165 -VNII--GLEDMKALGV-W-RSYVD-----WYRRDQTSABGIRQYQATYYAGWRP 213
 Qy 222 GNHLFLIN-----AYNSRRDLIERPMGHTTV-TLQCPALLVG----- 260
 Db 214 -EPDRWYQMGWYRGKRESVANNSALTVDIMFTQPVVVELDRLQWPTLLISEKONTA 272
 Qy 261 ---DSSPAVDVNEGSKLDP-----TKTLLKMAQDGGPQISQPAKLAFAF 305
 Db 273 ICKDAAPA-----EUKAKLVAYQLCKDAARRIPQATLVEFPDIGHYQIAPERFHOA- 326
 Qy 306 KYFVQGMGYMP 316
 Db 327 --LLEGLOTQ 335

RESULT 6
 H83276
 probable lipase PA2949 [imported] - *Pseudomonas aeruginosa* (strain PA01)
 C/Species: *Pseudomonas aeruginosa*
 C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C/Accession: H83276
 R/Stover, C.K.; Plam, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,
 ; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A/Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho
 A/Reference number: A82950; MUID:20437337; PMID:10984043
 A/Accession: H83276
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-315 <STO>
 A/Cross-references: GB:AE004721; GB:AE004091; NID:99949041; PIDN:AA06337.1; GSPDB:GN001
 A/Experimental source: strain PA01
 C/Genetics:
 A/Gene: PA2949

Query Match 6.8%; Score 140; DB 2; Length 315;
 Best Local Similarity 22.9%; Pred. No. 0.00083;
 Matches 78; Conservative 49; Mismatches 116; Indels 98; Gaps 18;

Qy 10 LAEVPFLVEKGETTIGLQEPVQODIET--LHGSVHTLCGFPKGRPVILTYHDGM 67
 Db 27 LASVR-TYERG-LAG-LSHSVQVDNLEIAYLB-----GSK-NPTLLIHGGA 73
 Qy 68 ---NHKTCYNPLFNYEDMOBITOHFAVCHVAPGQODGASFPAGYMPMSDOLAEMLP 124
 Db 74 DKDWLRFARPL-----TERHVVALDLPGGDSKPKQASV--DVGTQAEVAVN 121

Qy 125 VLOQFGKLSITGCTGAGAYILTRPALNPEWEGVLIIN---VNPCAGGMDMAASKIS 181
 Db 122 PAALIGVRRLHLAGNSGSHIALYARHPEQVLSLALIONAGWMPARK----- 170
 Qy 182 GWTQALPDVVVSHLFGKEMQSNVEVHTYRQ-----HIVDMNPGLHFLI-- 228
 Db 171 -----SELFEDLERGENPLVY---RQEDFPQKLDPFVVOQPPPLPAPLRYLGE 216
 Qy 229 ---NAYNS-----RDLIERPMGHTTVTLQCPALLVGDSSPAVDVNEGSKL 276
 Db 217 RAVASAFNNAQIFEQLRQRYTLPLEPLP-----KIEPFTLLWGDRLDRLVDV-----SSI 266
 Qy 277 DPTKTLTK-----MADCGLPQISQPAKLAFAKYPVQGM 312
 Db 267 EWRPRLKPSVIMENGVHWPVERBETRQHYQATLDGV 307

RESULT 7
 BVECBH
 biotin biosynthesis protein biob - *Escherichia coli* (strain K-12)

C/Species: *Escherichia coli*
 C/Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 01-Mar-2002
 C/Accession: J00081; G65136; S06071
 R/O'Regan, M.; Gloeckler, R.; Bernard, S.; Ledoux, C.; Ohnawa, I.; Lemoine, Y.
 Nucleic Acids Res. 17, 8004, 1989
 A/Title: Nucleotide sequence of the biob gene of *Escherichia coli*.
 A/Reference number: J00081; MUID:90016899; PMID:2678009
 A/Accession: J00081
 A/Molecule type: DNA
 A/Residues: 1-256 <GB>
 A/Cross-references: GB:X15587; NID:941067; PIDN:CAA33612.1; PID:941068
 R/Blatner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
 A.; Rose, D.T.; Mau, B.; Sha, Y.
 Science 277, 1453-1462, 1997
 A/Title: The complete genome sequence of *Escherichia coli* K-12.
 A/Reference number: A64720; MUID:97426617; PMID:9278503
 A/Accession: G65136
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-256 <BLAT>
 A/Cross-references: GB:AE000417; GB:U00096; NID:92367220; PIDN:AACT6437.1; PID:G1789817;
 A/Experimental source: strain K-12, substrain M61655
 C/Comment: This protein is involved in an early unknown step of the biotin biosynthesis
 C/Genetics:
 A/Gene: biob
 A/Map position: 75 min
 A/Superfamily: biob protein
 C/Keywords: biotin biosynthesis

Query Match 6.7%; Score 138.5; DB 1; Length 256;
 Best Local Similarity 24.2%; Pred. No. 0.00082;
 Matches 69; Conservative 35; Mismatches 80; Indels 101; Gaps 14;

Qy 42 GSVHTLCGFPKGRPVILTYHDGMNKT--CYNPLFNYEDMOBITOHFAVCHVAPG- 98
 Db 12 GNVLIVL-----LHGWLNAEVRKID-----ELSSHFTLHVLDPGF 50
 Qy 99 ---QODGASFPAGYMPMSDOLAEMLPGLVLOQFGKLSITGCTGAGAYILTRPALNPE 155
 Db 51 GRSRGFGLS-----LADMAEAVLQAPDKA-IWLGWSLGLVLSQALTHPE 97
 Qy 156 MVEGLVILNVNFCAGGMDMAASK--ISGHTQALPDVVVSHLFGKEMQSNVEVHTYR 212
 Db 98 RVQALVTVASSPCPSABDEWPGIRPDVLAGFQOQLSD-----DFQTVB----- 141
 Qy 213 OHIVDMNPGLHFLINAYNSRRDLIERPMGHTTVTLQCPAL--LVGDSSPAVDVAV 270
 Db 142 -----RFLAQ--TMGETARQDARALKYVLAALPMEVVDV- 176
 Qy 271 ECKSKLDPKTLTKMAQDGGLPQISQPAKLAFAKYPVQGMGYM 315
 Db 177 --NGGLEILKTVDLRQ-----LQNVSWP-----FLRLYGYL 206

RESULT 8

AB1655
prolyl aminopeptidases homolog ltn1782 [imported] - *Listeria innocua* (strain Clip11262)
C:Species: *Listeria innocua*
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 06-Jan-2003
C:Accession: AB1655
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Biocker, D.; Jones, L.M.; Karsch, U.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Feihl, H. Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Møller, C.; Schlueter, T.; Simoes, N.; Tixeront, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria species*.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AB1655
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-275 <GLA>
A:Cross-references: GB:AL592022; PIDN:CA97013.1; PID:gl6414269; GSPDB:GN00178
A:Experimental source: strain Clip11262
C:Genetics:
A:Gene: ltn1782
C:Superfamily: tropinesterase

Query Match 6.7%; Score 138.5; DB 2; Length 275;

Best Local Similarity 21.4%; Pred. No. 0.00091;

Matches 63; Conservative 49; Mismatches 124; Indels 59; Gaps 11;

QY 45 HTVLCTGPKGNRPVILTYHDIGMNHKTCNPLEFNEY-MOEITQHFVAVCHVDAPGQODGA 103
DB 7 HYLINDISGEKFPVLMHGFQSSSET-----FQDISLKHKFSIIADDLGH--GK 57
QY 104 ASFPAGWYPSMDQLAEMLPVYLQOFGKLSIIGMGTGAGAYILTRPALNPEWVEGLVLI 163
DB 58 TSCPEKVARYSIENICDDILASILHQLKIDSCFVLGYSWGGVATFAAKYKELVGLVLI 117
QY 164 NVNPGAEQMD--MAASKISGWTQALPDMVYVSHLFGKEMQSNVAVHTYQHLYNDMNP 221
DB 118 SSSP--GLRDEKAPASRISA-DNRFLADTLDA-----DGLEPVAIWE----- 156
QY 222 GNHLHFLINAYN---SRRLDEIERPMPGTHVT-----LQCPAL 257
DB 157 -NLALFASQKQLPFLKRRIRERLAQNPHGLAKSLRGWGTGKQPSYENLADFTFPVLI 215
QY 258 VVGSSPAVDVAVGNSKLDPTKTTLLKMDQCGLPQISQPAKLAFAKRYVQGM 312
DB 216 ITGNIDKFEKFIAREMKQLLP-NSTHYTVVQEGHVAVYLEQPNIFSSQLIYMLEGI 269

RESULT 9

AB7548
hypothetical protein CC2411 [imported] - *Caulobacter crescentus*
C:Species: *Caulobacter crescentus*
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: AB7548
R:Nierman, W.C.; Fedaldy, T.V.; Paulsen, I.T.; Nelson, K.E.; Eissen, J.; Heidelberg, J. B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.U.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolor, n, J.; Brimacombe, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M. Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete genome sequence of *Caulobacter crescentus*.
A:Reference number: AB7249; MUID:21173698; PMID:11259647
A:Accession: AB7548
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-393 <STO>
A:Cross-references: GB:AB005673; NID:g13423950; PIDN:AAK24382.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC2411

Query Match 6.5%; Score 135.5; DB 2; Length 393;

Best Local Similarity 22.8%; Pred. No. 0.0027;

Matches 62; Conservative 47; Mismatches 112; Indels 51; Gaps 12;

QY 55 NRPVILTYHDIGMNHKTCNPLEFNEYDMQITQHFVAVCHVDAPGQODGASFPAGWYPS 114
DB 20 DKPVLINSLGCD-LSLHDPV-----TPLTDPFRRLRIDTRGH--GASDAPSGDY--S 69
QY 115 MDQLAEMLPVYLQOFGKLSIIGMGTGAGAYILTRPALNPEWVEGLVLIWVPCAEQMD 174
DB 70 LDILADVAVLMMDAGAAKKTICTGSLGMIAMALASRAPDRVAVLACTSPAMD--- 126
QY 175 MAASKISGWTQ-----ALPDMVYVSHLFGKEMQSNVAVHTYQHLYNDMNP 223
DB 127 -----SSWECORLAVIRAEGLSIVAEVMSRFSDDFRALHPVEVETVRAGML-AQNP 179
QY 224 LHLFVNVRRLDIERPMPGTHVTVLQCPALLVGDSPPA-----VDAVVEGNSKLD 277
DB 180 ---YCGGGAIRMDALDRLP---KIAPVTLVTSKQVAIPPEGHADRIY---AAMP 228
QY 278 PTKTTLKMDCCGLPQISQPAKLAFAKRYV 309
DB 229 GARNAVIEAH--LPSLEAPAFAGAVRGFL 257

RESULT 10

F91160
biotin biosynthesis protein BioH [imported] - *Escherichia coli* (strain O157:H7, substrain
C:Species: *Escherichia coli*
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: F91160
R:Hayaishi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genom
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: F91160
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-256 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA37677.1; PID:g13363728; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: EC84255
C:Superfamily: bioH protein

Query Match 6.5%; Score 134.5; DB 2; Length 256;

Best Local Similarity 23.9%; Pred. No. 0.0017;

Matches 68; Conservative 36; Mismatches 80; Indels 101; Gaps 14;

QY 42 GSVHVLCTGPKGNRPVILTYHDIGMNHKT--CYNPLEFNEYDMQETQHFVAVCHVDAPG- 98
DB 12 GNHVLVL-----LHMGGLNABVWRCID-----ELSSHFTLHLVDLPGF 50
QY 99 ---QODGASFPAGWYPSMDQLAEMLPVYLQOFGKLSIIGMGTGAGAYILTRPALNPE 155
DB 51 GRSRGFALS-----LAEMAEVALRQAPDKA-IWLQMSIGLVAISQIALTHPE 97
QY 156 MVEGLVLIWVPCAEQMDAAK---ISGWTQALPDMVYVSHLFGKEMQSNVAVHTYR 212
DB 98 RVQALVTVASSPCFSARDEWPGIKPVYLAFQOQLSD-----DFQRYE----- 141
QY 213 QHLYNDMNPQNLHFLINAYVSRRLDEIERPMPGTHVTVLQCPAL--LVVGDSSPAVDVAV 270
DB 142 -----RFLALQ--TMGTETARQDARLAKKTVLALPMPVVDVL- 176
QY 271 ECNSKLDPTKTTLLKMDCCGLPQISQPAKLAFAKRYVQGMWY 315
DB 177 --NGLEILTKVDLRP---LQNVMP-----FLRLXYL 206

RESULT 11

E86006
biotin biosynthesis protein BioH [imported] - *Escherichia coli* (strain O157:H7, substrain
C:Species: *Escherichia coli*

C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 27-Nov-2001
 C:Accession: E86006
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew
 Iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A>Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: E86006
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-256 <STO>
 A:Cross-references: GB:AE005174; MID:g12518042; PIDN:AAG58513.1; GSPDB:GN00145; UWGP:247
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: bioH
 C:Superfamily: bioH protein

Query Match 6.5%; Score 134.5; DB 2; Length 256;
 Best Local Similarity 23.9%; Pred. No. 0.0017;
 Matches 68; Conservative 36; Mismatches 80; Indels 101; Gaps 14;

QY 42 GSVHTLTCTPKGNRPVLTTHYDGMNKT--CYNPLRYNEDMOITQFAVCHVDAG- 98
 DB 12 GNVHLVLT-----LHGMGLNABVWRCID-----EELSHFTLHVLDPGF 50
 QY 99 ---QODGAAPGAYNPSMDQLAEMLPGLVLOQFGKSLIGMGTAAGAYILTRPALNPE 155
 DB 51 GRSRFGALIS-----LAMAENVLKQADPKA-ILKMSLGLVLSQILTRPE 97
 QY 156 MVEGVLVLTINVPCEAGMDMAASK--ISGWTQALPDVNVSHLFKEEMQSNVEVHTYR 212
 DB 98 RVQALVTVAASPCFARDEMPKIDVLAAGFQQLSD-----DFQRIVE----- 141
 QY 213 QHIVDMNPGLHLFINAVNSRDLERPMFGTHTVTLQCAL--LVYGDSSPAVDVAV 270
 DB 142 -----RFLALQ--TWGTETARQDARALKKTVLALPMPEDVL- 176
 QY 271 ECNSKLDPTKTTLLKMAOCGLPOLISOPAKLAFAKVFVQMGVY 315
 DB 177 --NGLEILKTYDLRLP-----LQNVPM-----FLRLYGL 206

RESULT 12
 AB1284
 prolyl aminopeptidases homolog lmo1674 [imported] - listeria monocytogenes (strain EGD-e)
 C:Species: listeria monocytogenes
 C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 06-Jan-2003
 C:Accession: AB1284
 R:Glaesner, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bioecker
 ; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entlian, K.D.; Fsihi, H.
 Science 294, 849-852, 2001
 A:Authors: Kretz, T.; Kuhn, M.; Kunat, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
 ok, C.; Schlueter, T.; Simeos, N.; Tietzer, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,
 A>Title: Comparative genomes of listeria species.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AB1284
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-275 <GLA>
 A:Cross-references: GB:NC_003210; PIDN:CAC99752.1; PID:g16411110; GSPDB:GN00177
 A:Experimental source: strain EGD-e
 C:Genetics:
 A:Gene: lmo1674
 C:Superfamily: tropinesterase

Query Match 6.2%; Score 128.5; DB 2; Length 275;
 Best Local Similarity 20.3%; Pred. No. 0.0061;
 Matches 59; Conservative 48; Mismatches 129; Indels 55; Gaps 11;

QY 45 HTTLCGTPKGNRPVLTTHYDGMNKTCTNPLFNYD--MOEITQFAVCHVDAPGQODGA 103
 DB 7 HVHLLTAISGEKRALMLHGFTGTSET-----FQDSISGLKRFNIIAPDLGH--GN 57

QY 104 ASPGAYVPSMDQLAEMLPGLVLOQFGKSLIGMGTAAGAYILTRPALNPEVGLVLI 163
 DB 58 TASBEELSSYTMENICEDLAGIHLQLVNSRCPVLYGMSGVATAPATFPKRRLGLIV 117
 QY 164 NVNCEAGMDMAASKISGWTQALPDVNVSHLFKEEMQSNVEVHTYRQHI VDNMNGN 223
 DB 118 SSSPGLF--QEDIRASRVEA--DNRLADWI-----ELEGIVPVVDYWE-----N 157
 QY 224 LHLF-----INAVNSRDLERPMFGTHTVTL-----QC-----PALLIV 259
 DB 158 LALPASQKVSPEMKRIRSRSLQNSHGLMSLRGKGTGKQPSYNNCLANFTFPVLLIT 217
 QY 260 GDSSPAVDAYVECNKLDPTKTTLLKMAOCGLPOLISOPAKLAFAKVFVQ 310
 DB 218 GALDEKFEKIAQENHQLLP--NSTVSIQEGHVAVYLEQPNSSFSQLNWL 267

RESULT 13
 CB2692
 biotin biosynthesis protein xfl356 [imported] - xylella fastidiosa (strain 9a5c)
 C:Species: xylella fastidiosa
 C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: CB2692
 R:anonymous, The xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
 Nature 406, 151-157, 2000
 A>Title: The genome sequence of the plant pathogen xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717; PMID:10910337
 A>Note: for a complete list of authors see reference number A59328 below
 A:Accession: CB2692
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-255 <STM>
 A:Cross-references: GB:AE003967; GB:AE003849; MID:g9106347; PIDN:AAE84165.1; GSPDB:GN001
 A:Experimental source: strain 9a5c
 R:Simson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carrero, D.M.; Carre, H
 as-Neto, E.; Docena, C.; El-Dorriy, H.; Fedincant, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000
 A:Authors: Ferreira, V.C.A.; Petro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohm
 U.D.; Junqueira, M.L.; Kemper, E.L.; Kitzajima, J.P.; Kitzajima, J.B.; Kuramae, E.E.; Laig
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.B.; de Sa, R.G.; Santelli, R.V.; Sawasak
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, F.A.; da Silva Jr., W.A.; da Silveir
 M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: xfl356
 C:Superfamily: bioH protein

Query Match 6.2%; Score 127.5; DB 2; Length 255;
 Best Local Similarity 22.2%; Pred. No. 0.0066;
 Matches 64; Conservative 43; Mismatches 106; Indels 75; Gaps 15;

QY 54 GNRPVILTTHYDGMNKTCTNPLFNYEDMOEITQFAVCHVDAPGQODGAAPGAYVY 113
 DB 7 GYCPALVLIHGMAN-HSGVFAPL-----VEQLRAHHTLVLVDLPGH-----GYVHT 51
 QY 114 SMDQLAEMLPGLVLOQFGKLS--IIGMGTAAGAYILTRPALNPEVGLVLTINVPCEAG 171
 DB 52 TLTPLA--LPHVVAIAAATPPAAVWLMSLGLGALFAHAAATLPQ--VGLILMLAATPCVVR 108
 QY 172 WMDNA-ASKISGWTQALPDVNVSHLFKEEMQSNVEVHTYRQHI VDNMNGN 220
 DB 109 REDWPHAVVSIFFQ-----FAEDLKQNTERTINRLADLTGSTIAQSELR--- 155
 QY 221 PGNLHLFINAVNSRDLERPMFGT-----HT-----VTLCALLVGDSS 263
 DB 156 --QKQILNA-----NHTENTATLQAGLELALHTDLRRRAVIDLTLPSPMIAGORD 203

Qy 264 PAVDAV-VECNKSLDPT-KTTLKMAADCGGLPQISQPAKLAFAKYPV 309
 Db 204 RLVPAASIAHATALAPSGQTELLITGGGHAFLSHANQMTRALLQHF1 251

RESULT 14

AF0016
 Probable biotin biosynthesis protein bioH (imported) - Yersinia pestis (strain CO92)
 C:Species: Yersinia pestis
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
 A:Accession: AF0016
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarrag, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Dougan, G.; 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, Nature 413, 523-527, 2001
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: AF0016
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-258 <KOR>
 A:Cross-references: GB:AL590842; PIDN:CAC88992.1; PID:G15978234; GSPDB:GN00175
 C:Genetics:
 A:Gene: bioH
 C:Superfamily: bioH protein

Query Match 6.0%; Score 125; DB 2; Length 258;
 Best Local Similarity 23.1%; Pred. No. 0.011; Mismatches 112; Indels 64; Gaps 16;
 Matches 66; Conservative 44;

Qy 49 CGTPKGRPVILTYHDIGMNHKT--CYNPLFNYEDMOEITQHFAVCHVDAPG---QDDG 102
 Db 9 CG--EGDCDLVL-LHGMGLNAGVNHCI-----IDRLAPHFRLHVDLPYGRSGQDYG 57
 Qy 103 AASFPAGMTYPSMDOLAEMLRGVLDQFGKSLIGKGTGAGAYILTRFALNPNEMVEGLVL 162
 Db 58 AMS-----LADMAERVAQ--APQALWLGSMGGLVVASQIALSQPECVRGLIT 104
 Qy 163 INVNPACGEMDMWMAASK--ISGWTQALPD-----MVVSHLFGEKEMQSNVEVYHT- 210
 Db 105 VSSSPCFARDEWPGIKFEVLAFGPHQSLSDPHRTVERFLALQTLGTSSRQDARLKSV 164
 Qy 211 YRQHIIVDMN--PGNLHLFINAYNRRDLERPMPTGTHVTTLQCPALLVVGDSPPAVDA 268
 Db 165 VLQHQMPDVEVLTGLAIL-----RTADLRITLAG-----FTLPFRVYVGH---LDS 208
 Qy 269 VV--ECNKLDPPT--KTTLKMAADCGGLPQISQPAKLAFAKYPVQ 310
 Db 209 LVPRKVASLLDSAMPQTSVVMGAAHAFFISHPNDFAKILNFAE 254

RESULT 15

AE0997
 Probable biotin biosynthesis protein STY4287 (imported) - Salmonella enterica subsp. ent
 C:Species: Salmonella enterica subsp. enterica serovar Typh
 A:Note: this species has also been called Salmonella typh
 C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
 C:Accession: AE0997
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher, th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P. Nature 413, 848-852, 2001
 A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov A:Reference number: AB0502; MUID:21534947; PMID:11677608
 A:Accession: AE0997
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-256 <PAR>
 A:Cross-references: GB:AL513382; PIDN:CAD08105.1; PID:G16505084; GSPDB:GN00176
 C:Genetics:
 A:Gene: STY4287
 C:Superfamily: bioH protein

Query Match 5.9%; Score 123; DB 2; Length 256;
 Best Local Similarity 23.3%; Pred. No. 0.015; Mismatches 100; Indels 86; Gaps 14;
 Matches 67; Conservative 35;

Qy 53 KGRNPVILTYHDIGMNHKT--CYNPLFNYEDMOEITQHFAVCHVDAPGQDDGAASPPAGY 110
 Db 11 EGNCHLVL-LHGMGLNAGVNHCI-----EELGSHFTLHVLDPG--YGRSSGFGAMT 60
 Qy 111 MYPSMDOLAEMLRGVLDQFGKSLIGKGTGAGAYILTRFALNPNEMVEGLVLNVNCC-- 168
 Db 61 LBEXTAQVAAKNAP-----DQALWLGSLGGLVASQMALTPHERVQALVTVAASSPCFS 112
 Qy 169 -AEGMDWMAASKISGWTQALPDVNVSHLFGEKEMQSNVE-----VHATYRQHIIVDM 219
 Db 113 ARBWPQIKREILIGFPQQQLSD-----DPQRTVERFLALQTLGTETARQDA----- 158
 Qy 220 NPGNLHLFINAYNRRDLERPMPTGTH-----TVTLQ-----CPALLVVGDS 262
 Db 159 -----RTLSKVLAQMPDVEVLNGLLEIKTVDLREALKNVMPFLRLYG-- 204
 Qy 263 SPAVDAYECNKLDPPTKTTLK-----MADCGGLPQISQPAKLAFA 304
 Db 205 -YLDGLVP--RKIAPLDLTPHSTSQIMAKAHAFFISHPAFCA 248

Search completed: January 29, 2004, 23:29:51
 Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 23:25:08 ; Search time 17 seconds
(without alignments)
1089.913 Million cell updates/sec

Title: US-10-057-832-2
Perfect score: 2072
Sequence: 1 MSRMQVDVLAEVKPLVEKG.....TPNSGAAGNAGPKMEVSC 394

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues
Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2072	100.0	394	1 NDR1_HUMAN	Q92597 homo sapien
2	1960	94.6	394	1 NDR1_MOUSE	Q62433 mus musculu
3	1254	60.5	375	1 NDR3_HUMAN	Q9UGV2 homo sapien
4	1244	60.0	375	1 NDR3_MOUSE	Q9UGY9 mus musculu
5	1117	53.9	352	1 NDR4_HUMAN	Q9ALP0 homo sapien
6	1108	53.5	352	1 NDR4_RAT	Q9Z219 rattus norv
7	1049	50.6	371	1 NDR2_HUMAN	Q9UN36 homo sapien
8	1038	50.1	371	1 NDR2_MOUSE	Q9UGY0 mus musculu
9	351.5	17.0	352	1 SR21_HELAN	O23969 helianthus
10	347	16.7	325	1 VPJ1_CAERL	O02485 caenorhabdi
11	138.5	6.7	256	1 B10H_ECOLI	P13001 escherichia
12	107.5	5.2	1808	1 TENA_CHICK	P10039 gallus galli
13	99.5	4.8	331	1 ILVC_SYNEL	Q86G90 synecococc
14	99.5	4.8	543	1 MIG1_KLJMA	P52288 kluyveromyc
15	99	4.8	523	1 DHE4_CHLSO	P28998 chlorella s
16	98	4.7	966	1 CAPP_MEDSA	O02735 medicago sa
17	97.5	4.7	1353	1 VGL2_CVHOC	P36334 human coron
18	95	4.6	1159	1 RPOC_FORCN	O33431 porphyromon
19	94.5	4.6	750	1 YKS7_YEAST	P34231 saccharomyc
20	94.5	4.6	1242	1 IRS1_HUMAN	P33568 homo sapien
21	93.5	4.5	294	1 DEH1_MORSP	O01398 moraxella s
22	93.5	4.5	530	1 INR2_BOVIN	O95141 bos taurus
23	93.5	4.5	575	1 HEMA_SEBUNZ	P04853 sendai viru
24	93.5	4.5	1486	1 MUXB_ECOLI	P22523 escherichia
25	92.5	4.4	1850	1 GLPT_HAEIN	P02845 gallus galli
26	92	4.4	1850	1 VIT2_CHICK	P06863 sendai viru
27	91.5	4.4	575	1 HEMA_SENDJ	O15818 dictyosteli
28	91.5	4.4	1321	1 IFX3_DICDI	P00632 actinobact
29	90.5	4.4	266	1 ELM2_PACIC	Q57427 haemophilus
30	90.5	4.4	287	1 Y193_HAEIN	O16656 homo sapien
31	90	4.3	238	1 SFR7_HUMAN	Q00238 rattus norv
32	89.5	4.3	545	1 ICA1_RAT	P27562 sendai viru
33	89.5	4.3	575	1 HEMA_SEBNS	

ALIGNMENTS

RESULT 1	ID	NDRI_HUMAN	STANDARD:	PRT:	394 AA.
AC	Q92597	015207	Q9NVR6	Q9UK29	
DT	15-JUL-1999	(Rel. 38, Last sequence update)			
DT	15-JUL-1999	(Rel. 42, Last annotation update)			
DE	NDRI1 protein (N-myc downstream regulated gene 1 protein)				
DE	(Differentiation-related gene 1 protein) (DRG1) (Reducing agents and tunicamycin-responsive protein) (RTP) (Nickel-specific induction protein Cap43) (Ric42).				
GN	NDRI1 OR RTP OR DRG1 OR CAP43.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
CC	Mammalia; Eutazoa; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
NP	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Umbilical vein endothelial cells;				
RX	MEDLINE=97094664; PubMed=8939898;				
RA	Kokame K., Kato H., Miyata T.;				
RT	"Homocysteine-responsive genes in vascular endothelial cells identified by differential display analysis. GRP78/BiP and novel genes.";				
RL	J. Biol. Chem. 271:29659-29665(1996).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=97395543; PubMed=9251681;				
RA	van Belzen N., Dinjens W.N.M., Vliesveld M.P.G., Groen N.A.,				
RT	Boeman F.T.;				
RL	"A novel gene which is up-regulated during colon epithelial cell differentiation and down-regulated in colorectal neoplasms.";				
RN	Lab. Invest. 77:85-92(1997).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Lung;				
RX	MEDLINE=98266801; PubMed=9605764;				
RA	Zhou D., Salnikow K., Costa M.;				
RT	"Cap43, a novel gene, specifically induced by Ni2+ compounds.";				
RL	Cancer Res. 58:2182-2189(1998).				
RN	[4]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=99326165; PubMed=10395947;				
RA	Pignatelli D., Joulia D., Balaguer P., Basset A., Marti J., Comnes T.;				
RT	"Differential expression of the RTP/DRG1/NDR1 gene product in proliferating and growth arrested cells.";				
RL	Biochim. Biophys. Acta 1450:364-373(1999).				
RN	[5]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Kidney;				
RX	MEDLINE=22388257; PubMed=12477932;				
RA	Strauberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,				
RT	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,				
RL	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,				
RP	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,				

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stuplinter M., Soares M.B., Bonaldo M.F., Cacaiani T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Cantini P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gnatatue P.H.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Schermer A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RN [6]
 RP SEQUENCE OF 1-198 FROM N.A.
 RA Blechschmidt K., Kalaydjieva L., Goodman R., Gresham D., Baas F.,
 RA Jonge R.D., Schilabel M., Schattevoy R., Baumgart C., Menzel U.,
 RA Rosenthal A.;
 RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
 RN [7]
 RP SEQUENCE OF 1-21 FROM N.A.
 RC TISSUE=Brain;
 RA Angelicheva D., Kalaydjieva L.;
 RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
 RN [8]
 RP FUNCTION.
 RX MEDLINE=99438066; PubMed=9766676;
 RA Kudistean S.K., Arizti P., Reimer C.L., Sugrue M.W., Aaronson S.A.,
 RA Lee S.W.;
 RT "Inhibition of tumor cell growth by RTP/rict4 and its responsiveness
 to p53 and DNA damage.";
 RL Cancer Res. 58:4439-4444 (1998).
 CC -1- FUNCTION: MAY HAVE A GROWTH INHIBITORY ROLE.
 CC -1- SUBCELLULAR LOCATION: WHEREAS IN PROSTATE EPITHELIUM AND PLACENTAL
 CC CHORION IT IS LOCATED IN BOTH THE CYTOPLASM AND THE NUCLEUS,
 CC NUCLEAR STAINING IS NOT OBSERVED IN COLON EPITHELIUM CELLS.
 CC INSTEAD ITS LOCALIZATION CHANGES FROM THE CYTOPLASM TO THE PLASMA
 CC MEMBRANE DURING DIFFERENTIATION OF COLON CARCINOMA CELL LINES IN
 CC VITRO.
 CC -1- TISSUE SPECIFICITY: UBIQUITOUS; EXPRESSED MOST PROMINENTLY IN
 CC PLACENTAL MEMBRANES AND PROSTATE, KIDNEY, SMALL INTESTINE, AND
 CC OVARY TISSUES. REDUCED EXPRESSION IN ADENOCARCINOMAS COMPARED TO
 CC NORMAL TISSUES. IN COLON, PROSTATE AND PLACENTAL MEMBRANES, THE
 CC CELLS THAT BORDER THE LUMEN SHOW THE HIGHEST EXPRESSION.
 CC -1- INDUCTION: BY HOMOCYSTEINE, 2-MERCAPTOETHANOL, TUNICAMYCIN IN
 CC ENDOTHELIAL CELLS. INDUCED APPROXIMATELY 20-FOLD DURING IN VITRO
 CC DIFFERENTIATION OF THE COLON CARCINOMA CELL LINES HT29-D4 AND
 CC CACO-2. INDUCED BY NICKEL COMPOUNDS IN ALL TESTED CELL LINES. THE
 CC PRIMARY SIGNAL FOR ITS INDUCTION IS AN ELEVATION OF FREE
 CC INTRACELLULAR CALCIUM ION CAUSED BY NICKEL ION EXPOSURE. OKADAIC
 CC ACID, A SERINE/THREONINE PHOSPHATASE INHIBITOR, INDUCED ITS
 CC EXPRESSION MORE RAPIDLY AND MORE EFFICIENTLY THAN NICKEL.
 CC -1- SIMILARITY: BELONGS TO THE NDRG FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 CC EMBL; D87953; BA013505.1; -;
 DR EMBL; X92845; CA63430.1; -;
 DR EMBL; AF004162; AAC3419.1; -;
 DR EMBL; AF186190; -; NOT ANNOTATED_CDS.
 DR EMBL; BC003175; AA03175.1; -;
 DR EMBL; AF203080; AAF1305.1; -;
 DR GenBank; HGNC:7679; NDRG1.
 DR MIM; 605262; -;

DR GO; GO:0008181; F: tumor suppressor; TAS.
 DR GO; GO:0009412; P: response to heavy metal; TAS.
 DR InterPro; IPR004142; Ndr.
 DR InterPro; IPR003739; Ser_estre_site.
 DR Pfam; PF03096; Ndr; 1.
 KW Nuclear protein; Repeat.
 FT DOMAIN 339 368
 FT REPEAT 339 348
 FT REPEAT 349 358
 FT REPEAT 359 368
 FT CONFLICT 145 145
 FT SEQUENCE 394 AA; 42835 MW; 4C816B9C85B3756F CRC64;
 SQ
 Query Match 100.0%; Score 2072; DB 1; Length 394;
 Best Local Similarity 100.0%; Pred. No. 4.5e-162;
 Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSRMQDQVDAEYKPLVEKGGITGLQEPDVOEODIETLHGSVHTLCGTPKGNRPVIL 60
 DB 1 MSRMQDQVDAEYKPLVEKGGITGLQEPDVOEODIETLHGSVHTLCGTPKGNRPVIL 60
 QY 61 TYHDIGNNHKTCTNPLFNIEDMOEITDPAVCHVDAPGQDGAASPAGTYRSMQLA 120
 DB 61 TYHDIGNNHKTCTNPLFNIEDMOEITDPAVCHVDAPGQDGAASPAGTYRSMQLA 120
 QY 121 MLPGVLOQFGLKSLIGTGAGAYILTRPALNPNWEGVILNVNFCAGMDMAASKI 180
 DB 121 MLPGVLOQFGLKSLIGTGAGAYILTRPALNPNWEGVILNVNFCAGMDMAASKI 180
 QY 181 SGWTOALPDWVSHLPEKEMQSNVEVHTYRQHIYVNDMNGVHLFINAYSRRDLEIE 240
 DB 181 SGWTOALPDWVSHLPEKEMQSNVEVHTYRQHIYVNDMNGVHLFINAYSRRDLEIE 240
 QY 241 RPMPGHTHTVLQCPALIVVDSSPAVDVVECKSKLDPTTTLLKPADCGGLPQISPAK 300
 DB 241 RPMPGHTHTVLQCPALIVVDSSPAVDVVECKSKLDPTTTLLKPADCGGLPQISPAK 300
 QY 301 LAEAFKTFVQGMGMPASMTRLMRSTRASGSSTSLDGTGRSHTSEGRSHTSEGT 360
 DB 301 LAEAFKTFVQGMGMPASMTRLMRSTRASGSSTSLDGTGRSHTSEGRSHTSEGT 360
 QY 361 RSRSHTEGAHLDTTPNSGAAGNSAGPKSMVEVC 394
 DB 361 RSRSHTEGAHLDTTPNSGAAGNSAGPKSMVEVC 394
 RESULT 2
 NDRL MOUSE STANDARD; PRT; 394 AA.
 ID NDRL MOUSE
 AC 062433; P97862;
 DT 15-JUN-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE NDRL1 protein (N-myc downstream regulated gene 1 protein) (Protein
 DE Ndrl1).
 GN NDRL1 OR NDRL OR TDS.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99310663; PubMed=10381566;
 RA Shimono A., Okuda T., Kondoh H.;
 RT "N-myc-dependent repression of ndrl, a gene identified by direct
 RT subtraction of whole mouse embryo cDNAs between wild type and N-myc
 RT mutant.";
 RL Mech. Dev. 83:39-52 (1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Hydridoma;
 RX MEDLINE=97289706; PubMed=9144177;

Db 30 LIAEFKCYFGQGWGMPASMTLRMSRPAAGSSVTLSLEGTRSRHTSEGPRSRHTSEGS 3664

Qy 361 RSRHTSEGAHDTTPNSGAGNAGKPPKSMVSC 394

Db 361 RSRHTSEGARLINTPNSGATGNNAGPKRSMVSC 394

RESULT 3

ID NDR3_HUMAN STANDARD; PRT; 375 AA.

AC G09UGV2; Q96PL8; Q9BX7; Q9H3N7; Q9H411; Q9HJ6;.

DT 16-OCT-2001 (Rel. 40, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE NDRG3 protein.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.

RX MEDLINE=21251006; PubMed=11352569;

RA Zhou R.-H., Kokame K., Tsukamoto Y., Yutani C., Kato H., Miyata T.;

RT "Characterization of the human NDRG3 gene family: a newly identified member, NDRGA4, is specifically expressed in brain and heart.";

RL Genomics 73:86-97(2001).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORM 2).

RA Mao Y., Xie Y., Zhou Z., Zhao W., Zhao S., Wang W., Huang Y., Wang S., Tang R., Chen X., Wu C.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A. (ISOFORM 1).

RC TISSUE=Splice;

RX MEDLINE=21933870; PubMed=11936845;

RA Qu X., Zhai Y., Wei H., Zhang C., Xing G., Yu Y., He F.;

RT "Characterization and expression of three novel differentiation-related genes belonging to the human NDRG gene family.";

RL Mol. Cell. Biochem. 229:35-44(2002).

RN [4]

RP SEQUENCE FROM N.A. (ISOFORM 3).

RC TISSUE=Placenta;

RA Isogai T., Ota T., Hayashi K., Sugiyama T., Oseuki T., Suzuki Y., Nishikawa T., Nagai K., Sugano S., Shitatori A., Sudo H., Magesawa M., Hosofiri T., Kaku Y., Kodaira H., Kondo H., Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S., Yamamoto S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K., Yamanote S., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y., Niinomiya K., Iwayanagi T.;

RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

RN [5]

RP SEQUENCE FROM N.A.

RX MEDLINE=21638749; PubMed=11780052;

RA Delidakis P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R., Jones M., Stavriliadis G., Almeida J.P., Babbage A.K., Baggaley C.L., Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M., Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J., Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P., Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M., Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R., Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M., Ellington A.G., Frankland J.A., Fraser A., French L., Garner P., Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E., Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J., Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D., Kay M.P., Kimberley A.M., King A., Knights K., Laird G.K., Lawlor S., Levesajeth M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D., Marsh V.L., Martin S.L., McConnachie L.J., McClay G., McMurray A.A., Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T., Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.T., Phillimore B.U.C.T., Pratchalingam S.R., Plumb R.W., Ramsay H.,

RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,
 RA Skane C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.M., Symamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Williams S.A.,
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.!

RT "The DNA sequence and comparative analysis of human chromosome 20.",
 RL Nature 414:865-871(2001).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=3;
 CC Name=1;
 CC IsoId=Q9UGV2-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9UGV2-2; Sequence=VSP_003419;
 CC Name=3;
 CC IsoId=Q9UGV2-3; Sequence=VSP_003420;
 CC -1- TISSUE SPECIFICITY: Ubiquitous. Highly expressed in brain.
 CC -1- SIMILARITY: BELONGS TO THE NDRG FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AB044943; BAB20067.1; -
 DR EMBL; AF251054; AAK34944.1; -
 DR EMBL; AF08609; AAL08807.1; -
 DR EMBL; AK023618; BAB14620.1; -
 DR EMBL; AL132768; CAC10098.2; -
 DR EMBL; AL031662; CAB56525.1; ALT_SEQ.
 DR Genew; HGNC:14462; NDRG3.
 DR MIM; 605273; -
 DR GO; GO:0005737; C:cytoplasm; IDA.
 DR GO; GO:0030154; P:cell differentiation; NAS.
 DR GO; GO:0030308; P:negative regulation of cell growth; NAS.
 DR GO; GO:0007283; P:spermatogenesis; NAS.
 DR InterPro; IPR004142; Ndr.
 DR InterPro; IPR000379; Ser_estrs_site.
 DR Pfam; PF03096; Ndr; 1.
 DR KX Alternative splicing.
 DR VARSPLIC 20 31 Missing (in isoform 2).
 FT VARSPLIC 47 135 Missing (in isoform 3).
 FT VARSPLIC 47 135 Missing (in isoform 3).
 FT CONFLICT 32 32 E -> G (IN REF. 3).
 FT CONFLICT 122 122 P -> S (IN REF. 2).
 SQ SEQUENCE 375 AA; 41408 MW; 237B6DF43AB7F1E CRC64;

Query Match 60.5%; Score 1254; DB 1; Length 375;
 Best Local Similarity 61.8%; Pred. No. 3.6e-95;
 Matches 244; Conservative 58; Mismatches 67; Indels 26; Gaps 6;

4 EMOVDVLAIVKPLVKEGRTTGLDGFVQEDIDETLHGSVHTLCTGPKGNRPVILTYH 63
 3 ELQDVLTETIKPLL-NDKNGTRNPFQDFDQCEHDIETPHGMVHTLTLGLPKGNRPVILTYH 61
 64 DIGMNHKTCYNPLFENYEDMOEITQHPAVCHVAPGQDGAASFPAGYMPSPMDLAEMLP 123
 62 DIGLNHKSCTFAFNFEEDMOEITQHPAVCHVAPGQDGAASFPAGYMPSPMDLAEMLP 121
 124 GYLQOFGKLSITIGMGTAGAYITLTPALNNPEVVEGLVINVPDCAEGMDMAASKISGL 181
 122 PVLTHLSIKSTIIGVGAGAYITLTPALNNPEVVEGLVINVPDCAEGMDMAASKISGL 181
 184 TQALPDVNVSHLFEKEMQSNVEVHTYRQHTIVDMMNGULHFLINVNSRRDLEIRPM 243
 182 TTNVVDIILAHNFQDEIQANLQITRMHIAQDINQNLQFLNSYNRGRDLEIRPI 241
 244 PGTH--TVTLQCPALLVGDSPAVDAVVECNKSLDPTXTTLTKMADCGGLPQISQPAK 300

DB 242 LGQDNKSKTLKCTSTLLVVDNSPAVAVVECNRLNPIINTLLTKADCGGLPQVVGPK 301
 QY 301 LAEFKFKVGMGMPSASTRLMRST-ASGSVSUJSDTTRSRSHSTSEGTTRSRSHSTSE 359
 DB 302 LTFKFKFLGGMGYIPASWTRLARSRHTSTSSLSGSGSPFSRSVLT-----SNQSD 354

QY 360 TRSRSHSTSEGAHLDITPNSGAAGNSAGPKSMETSC 394
 DB 355 TQESCEP-----DVLDRH-----QTEVESC 375

RESULT 4
 NDR3 MOUSE
 ID NDR3 MOUSE STANDARD; PRT; 375 AA.
 AC Q9QYF9;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE NDRG3 protein (Ndr3 protein).
 GN NDRG3 OR NDR3.
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=20050077; PubMed=10581191;
 RA Okuda T., Kondoh H.;
 RT "Identification of new genes ndr2 and ndr3 which are related to
 RT Ndr1/RTF/Drg1 but show distinct tissue specificity and response to
 RT N-myc.";
 RL Biochem. Biophys. Res. Commun. 266:208-215 (1999).
 CC -1- TISSUE SPECIFICITY: THYMOS, NERVOUS SYSTEM, GUT AND KIDNEY.
 CC -1- DEVELOPMENTAL STAGE: ITS EXPRESSION IS ALREADY SIGNIFICANT AT 9.5
 CC DPC, COVERING THE ENTIRE EMBRYO EXCEPT THE HEART, AND IT SHOWS
 CC ONLY A SLIGHT INCREASE IN LATER DEVELOPMENTAL STAGES.
 CC -1- SIMILARITY: BELONGS TO THE NDRG FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AB033922; BAA58883.1; -
 DR MGI; MGI:1352499; Ndr3.
 DR GO; GO:0005737; C:cytoplasm; ISS.
 DR GO; GO:0030154; P:cell differentiation; ISS.
 DR GO; GO:0030308; P:negative regulation of cell growth; ISS.
 DR GO; GO:0007283; P:spermatogenesis; ISS.
 DR InterPro; IPR004142; Ndr.
 DR InterPro; IPR000379; Ser_estrs_site.
 DR Pfam; PF03096; Ndr; 1.
 SQ SEQUENCE 375 AA; 41555 MW; B33D3CC3B816A8A1 CRC64;

Query Match 60.0%; Score 1244; DB 1; Length 375;
 Best Local Similarity 61.0%; Pred. No. 2.4e-94;
 Matches 241; Conservative 58; Mismatches 70; Indels 26; Gaps 6;

4 EMOVDVLAIVKPLVKEGRTTGLDGFVQEDIDETLHGSVHTLCTGPKGNRPVILTYH 63
 3 ELQDVLTETIKPLL-NDKNGTRNPFQDFDQCEHDIETPHGMVHTLTLGLPKGNRPVILTYH 61
 64 DIGMNHKTCYNPLFENYEDMOEITQHPAVCHVAPGQDGAASFPAGYMPSPMDLAEMLP 123
 62 DIGLNHKSCTFAFNFEEDMOEITQHPAVCHVAPGQDGAASFPAGYMPSPMDLAEMLP 121
 124 GYLQOFGKLSITIGMGTAGAYITLTPALNNPEVVEGLVINVPDCAEGMDMAASKISGL 183


```

Db 122 PVLTLMSKSIIGICVAGAVILSRPALNHEPLVGLVINDPCAKMIDWASKLGF 181
Qy 184 TQALPDVTVSHLFGKEWQSVVHTYRHOIVNDMPGNLHFLINAVNSRDLIERPM 243
Db 182 TTNIVDIIIAHFGQEEIQANLNDLQYTRHIAQIINGENLQFLGSIINGRDLIERPI 241
Qy 244 PG---THTVTLQCPALLVGVSSPAVDVVEGNSKLDPTKTLTLMDACGLPQISQPAK 300
Db 242 LGQNDNRLLKTKCSTLLVGVGDSPAVEAVEGNSRLDPTINTLLMDACGLPQVQPGK 301
Qy 301 LAEAKYVQGMGVGPSASMTRLMKSRT-ASGSSVTSIDGTRSRKHTSEGRSRKHTSEG 359
Db 302 LTEAKYVQGMGVGPSASMTRLMKSRTSSISGSESPRSRST-----SNQSDG 354
Qy 360 TRSRHTSEGAHLDTTPNSGAAGNAGPKSMVESC 394
Db 355 TQESCESP-----DVLDRH-----QTMVEVC 375

RESULT 5
NDRA_HUMAN STANDARD: PRT: 352 AA.
AC Q9ULP0: Q96PL9; Q9GZM1; Q9GZN3; Q9GZX0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE NDRG4 protein (Brain development-related molecule 1) (Vascular smooth
DE muscle cell associated protein-8) (SMAP-8).
GN NDRG4 OR BDM1 OR KIAA1180.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.
RX MEDLINE=21251006; PubMed=11352569;
RA Zhou R.-H., Kokame K., Tsukamoto Y., Yutani C., Kato H., Miyata T.;
RT "Characterization of the human NDRG gene family: a newly identified
RT member, NDRG4, is specifically expressed in brain and heart.";
RL Genomics 73:66-97(2001).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Heart;
RA Nishimoto S., Tawara J., Aoki T., Toyoda H., Komurasaki T.;
RT "Molecular cloning and characterization of the human vascular smooth
RT muscle cell associated protein-8 (SMAP-8).";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 4).
RC TISSUE=Spleen;
RX MEDLINE=21931870; PubMed=11936845;
RA Qu X., Zhai Y., Wei H., Zhang C., Xing G., Yu Y., He F.;
RT "Characterization and expression of three novel
RT differentiation-related genes belong to the human NDRG gene family.";
RL Mol. Cell. Biochem. 229:35-44(2002).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Amygdala;
RX MEDLINE=21154917; PubMed=11230166;
RA Wiseman S., Weil B., Wellenreuther R., Gassenhuber J., Glasiel S.,
RA Ansoerg W., Boecker M., Bloecker H., Bauersachs S., Blum H.,
RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
RA Mewes H.-W., Ottenwaelder B., Obermajer B., Tampe J., Heubner D.,
RA Wandt R., Korn B., Klein M., Poustka A.;
RT "Towards a catalog of human genes and proteins: sequencing and
RT analysis of 500 novel complete protein coding human cDNAs.";
RL Genome Res. 11:422-435(2001).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=22158633; PubMed=12168954;
RA Nakajima D., Okazaki N., Yamakawa H., Kikuno R., Ohara O., Nagase T.;
RT "Construction of expression-ready cDNA clones for KIAA genes: manual

```

```

RT curation of 330 KIAA cDNA clones.";
RL DNA Res. 9:99-106(2002).
RN [6]
RP SEQUENCE OF 80-352 FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=20039618; PubMed=10574461;
RA Hirotsawa M., Nagase T., Ishikawa K.-I., Kikuno R., Nomura N.,
RA Ohara O.;
RT "Characterization of cDNA clones selected by the GenMark analysis
RT from size-fractionated cDNA libraries from human brain.";
RL DNA Res. 6:329-336(1999).
RN [7]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Muscle;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield Y.S.N., Krzywinski M.I., Skliska V., Smallegange D.B.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- FUNCTION: MAY PLAY A ROLE IN THE EARLY POSTNATAL DEVELOPMENT AND
CC -!- FUNCTION OF NEURONAL CELLS (BY SIMILARITY).
CC -!- ALTERNATIVE PRODUCTS:
CC Name=1; Synonyms=NDRG4-BVAR;
CC Name=2; Synonyms=NDRG4-B;
CC Name=3; Synonyms=NDRG4-H;
CC Name=4;
CC IsoId=G9ULP0-1; Sequence=VSP_003422;
CC IsoId=G9ULP0-2; Sequence=VSP_003421;
CC IsoId=G9ULP0-3; Sequence=VSP_003421, VSP_003422;
CC Name=4;
CC IsoId=G9ULP0-4; Sequence=VSP_003423;
CC -!- TISSUE SPECIFICITY: Expressed only in brain and heart. Isoforms 1
CC and 2 are only expressed in brain. Isoform 3 is expressed in both
CC heart and brain.
CC -!- SIMILARITY: BELONGS TO THE NDRG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; AB044947; BAB20071.1; -
DR EMBL; AB044947; BAB20072.1; -
DR EMBL; AB044947; BAB20073.1; -
DR EMBL; AB044944; BAB20068.1; -
DR EMBL; AB044945; BAB20069.1; -
DR EMBL; AB044946; BAB20070.1; -
DR EMBL; AB021172; BAB20288.1; -
DR EMBL; AF308608; BAB08806.1; -
DR EMBL; AL136584; CAB66519.1; -
DR EMBL; AB033006; BAB6494.2; ALT_INIT.
DR EMBL; BC011795; AAH11795.1; -
DR Genew; HGNC:14466; NDRG4.
DR GO; GO:0005737; C:cytoplasm; NAS.

```

```

DR GO; GO:0030154; P:cell differentiation; NAS.
DR GO; GO:0016049; P:cell growth; NAS.
DR GO; GO:0006950; P:response to stress; NAS.
DR InterPro; IPR004142; Ndr.
DR Pfam; PF03096; Ndr; 1.
DR Developmental protein; Alternative splicing.
FT DOMAIN 253
FT VARSPLIC 1
FT VARSPLIC 289 302
FT VARSPLIC 291 303
FT VARSPLIC 352 AA; 38459 MM; 4CFAFC7C820013B6 CRC64;
SQ SEQUENCE 352 AA; 38459 MM; 4CFAFC7C820013B6 CRC64;

Query Match 53.9%; Score 1117; DB 1; Length 352;
Best Local Similarity 57.2%; Pred. No. 5.4e-84;
Matches 214; Conservative 51; Mismatches 67; Indels 42; Gaps 3;

QY 34 EODIETLHSGVHTLCTGPKGNRPVILTYHDIGNNHKTVCNPLFNYEDMOETIQHFAVCH 93
DB 8 EHDIEPTGILHVIKRSRPGKGNRPALITVDGALNHLCEFTFNFEDMOETIKHFAVCH 67
QY 94 VDAFGQDGAASFPAGYMPMDQLAEMLPGLVQOGLKSIIGMGAGAYILTRFALNN 153
DB 68 VDAFGQDGAASQFPGYQFSPMEQLAANLPVVOHFGFKYIGIGVAGANVLAKFALIF 127
QY 154 PEMVEGLVILVNPACAGMDMAASKISGWTQALPDMVSHLFGKEKMOSEVAVHTYRQ 213
DB 128 PDLVEGLVILVNIIDNGKGMIDMAATKLSGLTSTLPDVTLSHLSFQSEELVNNTELVSQYRQ 187
QY 214 HLYVNDMPGNLHLFTIANYNSRDLIERPMPGHTVTLQCPALLVGDSSPAVDAYVECN 273
DB 188 QIGVNVQANLQLTWMNYSRRDLINPGTVPAKTLRCFVNLVGDNAFADGVVECN 247
QY 274 SKLDPRTKTLTKMADCGGLPOISQPAKLAFAKIFYVQMGY-----MPSASM 320
DB 248 SKLDPRTTTLTKMADSGGLPVQVTPGKLTFAKTFLOMGYIAHLKDRRLSGANVPASAM 307
QY 321 TRLMRSTASGSSVTSIDGTRRSHTSEGTSSHTSEGTSSHTSEGAHLDTTPNSGA 380
DB 308 TRLMRSTASLTSSASVDSGRPOACT-----HSESESGQGVNHT----- 347
QY 381 AGNSAGPKSMEVSC 394
DB 348 -----MEVSC 352

RESULT 6
NDR4 RAT STANDARD; PRT; 352 AA.
ID NDR4 RAT STANDARD; PRT; 352 AA.
AC Q922E9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE NDRG4 protein (Brain development-related molecule 1).
GN NDRG4 OR BDM1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99255674; PubMed=10320792;
RA Yamauchi Y., Hongo S., Nishinaka N., Ohashi T., Takahashi R.,
RA Takeda M.;
RT "Molecular cloning and characterization of a novel developmentally
RT regulated gene, Bdm1, showing predominant expression in postnatal rat
RT brain.";
RL Brain Res. Mol. Brain Res. 68:149-158(1999).

```

```

CC -1- FUNCTION: MAY PLAY A ROLE IN THE EARLY POSTNATAL DEVELOPMENT AND
CC FUNCTION OF NEURONAL CELLS.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE BRAIN AND HEART. HEAVILY IN
CC THE KIDNEY, MOST PROMINENTLY IN POSTNATAL BRAIN WHERE IT IS
CC EXPRESSED WIDELY IN THE OLFACTORY BULB, CEREBRAL CORTEX,
CC HIPPOCAMPUS, CEREBELLUM, THALAMUS, AND MEDULLA OBLONGATA.
CC -1- SIMILARITY: BELONGS TO THE NDRG FAMILY.
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL; AF045564; M202415.1; -.
DR InterPro; IPR004142; Ndr.
DR Pfam; PF03096; Ndr; 1.
DR Developmental protein.
FT DOMAIN 253
FT VARSPLIC 253
SQ SEQUENCE 352 AA; 38487 MM; EF3SD773B9D255D9 CRC64;

Query Match 53.5%; Score 1108; DB 1; Length 352;
Best Local Similarity 57.2%; Pred. No. 2.9e-83;
Matches 214; Conservative 50; Mismatches 68; Indels 42; Gaps 3;

QY 34 EODIETLHSGVHTLCTGPKGNRPVILTYHDIGNNHKTVCNPLFNYEDMOETIQHFAVCH 93
DB 8 EHDIEPTGILHVIKRSRPGKGNRPALITVDGALNHLCEFTFNFEDMOETIKHFAVCH 67
QY 94 VDAFGQDGAASFPAGYMPMDQLAEMLPGLVQOGLKSIIGMGAGAYILTRFALNN 153
DB 68 VDAFGQDGAASQFPGYQFSPMEQLAANLPVVOHFGFKYIGIGVAGANVLAKFALIF 127
QY 154 PEMVEGLVILVNPACAGMDMAASKISGWTQALPDMVSHLFGKEKMOSEVAVHTYRQ 213
DB 128 PDLVEGLVILVNIIDNGKGMIDMAATKLSGLTSTLPDVTLSHLSFQSEELVNNTELVSQYRQ 187
QY 214 HLYVNDMPGNLHLFTIANYNSRDLIERPMPGHTVTLQCPALLVGDSSPAVDAYVECN 273
DB 188 QISSVNVQANLQLTWMNYSRRDLINPGTVPAKTLRCFVNLVGDNAFADGVVECN 247
QY 274 SKLDPRTKTLTKMADCGGLPOISQPAKLAFAKIFYVQMGY-----MPSASM 320
DB 248 SKLDPRTTTLTKMADSGGLPVQVTPGKLTFAKTFLOMGYIAHLKDRRLSGANVPASAM 307
QY 321 TRLMRSTASGSSVTSIDGTRRSHTSEGTSSHTSEGTSSHTSEGAHLDTTPNSGA 380
DB 308 TRLMRSTASLTSSASVDSGRPOACT-----HSDSESGMGQVNH----- 347
QY 381 AGNSAGPKSMEVSC 394
DB 348 -----MEVSC 352

RESULT 7
NDR2 HUMAN STANDARD; PRT; 371 AA.
ID NDR2 HUMAN STANDARD; PRT; 371 AA.
AC Q9UN36; Q96FD3; Q96FT0; Q96GT0; Q96PN0; Q96OH5; Q9ULH2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-FEB-2003 (Rel. 42, Last annotation update)
DE NDRG2 protein (Syd1709613 protein).
GN NDRG2 OR SYD17 OR KIAA1248.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RA Deng Y.C., Yao L.B., Su C.Z., Lui X.P.;

```

RN Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Brain;
 RX MEDLINE=20039619; PubMed=10574462;
 RA Nagase T., Ishikawa K.-I., Kikuno R., Hirosewa M., Nomura N.,
 RA Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XV.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL Nucleic Acids Res. 27:337-345(1999).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Amygdala;
 RX MEDLINE=21154917; PubMed=1123016f;
 RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glasl S.,
 RA Amberg W., Boehr M., Bloecher H., Baerach S., Blum H.,
 RA Leuber J., Dueschhoeft A., Beyer A., Koehrer K., Strack N.,
 RA Lawers H.-W., Ottenwelder B., Obermair B., Tampe J., Heubner D.,
 RA Wambert R., Korn B., Klein M., Poustka A.;
 RT "Towards a catalog of human genes and proteins: sequencing and
 RT analysis of 500 novel complete protein coding human cDNAs.";
 RL Genome Res. 11:422-435(2001).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC MEDLINE=21251006; PubMed=11352569;
 RX Zhou R.-H., Kokame K., Tsukamoto Y., Yutani C., Kato H., Miyata T.;
 RA "Characterization of the human NDRG gene family: a newly identified
 RA member, NDRG4, is specifically expressed in brain and heart.";
 RL Genomics 73:86-97(2001).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC MEDLINE=21933870; PubMed=11936845;
 RX Qu X., Zhai Y., Wei H., Zhang C., Xing G., Yu Y., He F.;
 RA "Characterization and expression of three novel
 RA differentiation-related genes belong to the human NDRG gene family.";
 RL Mol. Cell. Biochem. 229:35-44(2002).
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Brain;
 RX Liu X., Deng Y., Yao L.;
 RT "Cloning and analysis of human genomic DNA sequence of N-myc
 RT downstream regulator 2.";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3).
 RC TISSUE=Brain;
 RX MEDLINE=22386257; PubMed=12477932;
 RA Straube R.L., Felting E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Butco K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heif F.,
 RA Datchenko L., Marsina K., Farmer A.S., Rubin G.M., Hong L.,
 RA Stepieton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loguailano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Raheij J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Bakkeney R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butcherfield J.S.N., Krzywnski M.I., Skalska U., Smallue D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- ALTERNATIVE PRODUCTS;
 CC Name=1;
 CC IsoId=09UN36-1; Sequence=Displayed;
 CC Name=2; Synonym=NDRG2var;

```

CC      IsoId=g9UN36-2; Sequence=VSP_003417;
CC      Name=3;
CC      IsoId=g9UN36-3; Sequence=VSP_003418;
CC      -1- SIMILARITY: BELONGS TO THE NDRG FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC      or send an email to license@ebi.ac.uk).
CC      -----
DR      EMBL; AF159092; AAD43131.2; -
DR      EMBL; AB033074; BAA86562.1; ALT_INIT.
DR      EMBL; AL136574; CAB66509.1; -
DR      EMBL; AF304051; AAL08624.1; -
DR      EMBL; AY028430; AAK50340.1; -
DR      EMBL; BC010458; AAH10458.1; -
DR      EMBL; BC011240; AAH11240.1; -
DR      Genew; HGNC:14460; NDRG2.
DR      MIM; 605272; -
DR      GO; GO:0005829; C:cytosol; NAS.
DR      GO; GO:0030154; P:cell differentiation; ISS.
DR      InterPro; IPR004142; Ndr.
DR      Pfam; PF03096; Ndr; 1.
KW      Alternative splicing.
FT      VASPLIC 26 39 Missing (in isoform 2).
FT      VASPLIC 262 272 Missing (in isoform 3).
FT      VASPLIC 262 272 Missing (in isoform 3).
FT      CONFLICT 48 48 G -> V (IN REF. 7).
FT      CONFLICT 54 54 V -> A (IN REF. 5).
FT      CONFLICT 172 172 D -> G (IN REF. 6).
FT      CONFLICT 250 250 Q -> N (IN REF. 5).
FT      CONFLICT 296 296 Q -> R (IN REF. 6).
FT      CONFLICT 306 306 A -> V (IN REF. 5).
SQ      SEQUENCE 371 AA; 40798 MW; 7B49B0B12BD3F595 CRC64;
Query March 50.6%; Score 1049; DB 1; Length 371;
Best Local Similarity 52.1%; Pred. No.2;le-78;
Matches 208; Conservative 64; Mismatches 89; Indels 38; Gaps 6;
QY      4 EMQDVLDAEYKPLV-----EKGETITGLQEFVQEDIDETLHSGSVHTLCTGPKGN 55
DB      3 ELGEOVITEKPLPLPGQTPKAKEALAEALRIILD-QGQTHSVETPYGVSFPTVYGTPEPK 61
QY      56 RPYLLTHHDIGMNHKTQYNPLFNVEDMQEITQHFVNCVDAFPGQDGAASFPAGTPTSM 115
DB      62 RPAITLTHDVLNLYKSCFQPLQFQEDMQEITQHFVNCVDAFPGMEBGAVPFPLQYQPSL 121
QY      116 DOLAEMLPGVLOQEGLSIIGMGAGAYILTRPALNPNPEVBEGLVILNVPCAEGNMD 175
DB      122 DQLDLMDLPCVLQYNSTIIGVGAGAYILARALNHPDVBGLVLINIDPAKGMNDW 181
QY      176 AASKISGWTALPDMVYSHLFGKEKMSQNVVHTYRQHTIVNDNPNGLHFLINAYNSRR 235
DB      182 AAHLITGLTSSIPMILGHLFSGEELSGNELLOKYNIIITHAPNDLNIELVYNSYNRR 241
QY      236 DLEIERPMGTHTVTLLOCPALLVYGSSPAVDAVECNSKLDPTKTTLLKKADCGSLPQI 295
DB      242 DLNFER-----GQDITTECPVMLVVGDAFHEBDVAVECNKSLDPTQTFLLKQADSGQPOL 297
QY      296 SOPKLAEARKYFQOGMGVMPASMTPLMSRTTASGSSVTSLSQSTRSRSHTSBETRSH 355
DB      298 TQPKLTLEARKYFQOGMGVMASSCMTRLSRTTASLSAASVDGNRSR----- 346
QY      356 TSEGRSRSHTSRGALHDTTPNSGAAGNSAGPYKMEVSC 394
DB      347 -----RTLSQSSBSG-----TLSSGPPGH-----TMEVSC 371

```

```

NDR2_MOUSE
ID_NDR2_MOUSE STANDARD; PRT; 371 AA.
AC_O90YG0.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE NDRG2 protein (Ndr2 protein).
GN NDRG2 OR NDR2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=20050077; PubMed=10581191;
RA Okuda T., Kondoh H.;
RT 'Identification of new genes ndr2 and ndr3 which are related to
RT Ndr1/RTP/Drp1 but show distinct tissue specificity and response to
RT N-myc.';
RL Biochem. Biophys. Res. Commun. 266:208-215(1999).
CC -!- TISSUE SPECIFICITY: HEART, NERVOUS SYSTEM, GUT AND KIDNEY.
CC -!- DEVELOPMENTAL STAGE: THE EXPRESSION IS QUITE LOW AT THE STAGE OF
CC 9.5 DPC, BUT BEGAN TO INCREASE AFTER 11.5 DPC.
CC -!- SIMILARITY: BELONGS TO THE NDRG FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB033921; BAA85882.1; -.
DR MGD; MGI:1352498; Ndr2.
DR InterPro; IPR004142; Ndr.
DR Pfam; PF03096; Ndr; 1.
SQ SEQUENCE 371 AA; 40789 MW; A5237C04278197B2 CRC64;

```

Query Match	Similarity	50.1%	Score 1038	DB 1	Length 371
Best Local	Similarity	52.1%	Pred. No. 1,7e-77		
Matches	208	Conservative	64	Mismatches	89
				Indels	38
				Gaps	7

Q	D	3	4	56	62	116	122	176	182	236	242	296	298	356	347
EMQDVLAEXPLV-----	ELQEQVETREKPLFGQTPRETKAEHLAARILLD-QGQTHSVETGYSTFVYVTPKPK	RPVLITTHDDIGMNHNTCYNPLFNEDMDQEIQHFALVCHVDAPGQDDGAASPAGMYRSM	EMQDVLAEXPLV-----	RPVLITTHDDIGMNHNTCYNPLFNEDMDQEIQHFALVCHVDAPGQDDGAASPAGMYRSM	RPALFTYHDVGLNYSFCQPLRFREDQMEIILQNFRAVHVADAPGMEEGAVPPLGQYPSL	DQLAEMLPGVILQOQFGLKSLIIGMGTGAGAYIILTRFPLNPNMEWEGVILINVPCEAGMMDW	DQLADMICILIQYLNFSITIIIGVGAGAYILSRVALNHPDVEGLVILINIDPNKAGMMDW	AASKISGCTQALPDWVNSHLRCKEEMQSNVEVHTYRQHIYVNDKMPGLHLEFINAYNSRR	AAHKLTGLTSSIPDMILIHGLFSQEBELGNSBELIQCKRGIIGHAPLENILEIYMNYSNNRR	DLEIRPMPGTHVTVLOCPALLVVDSSPAVDAYVEGNSKLDPTTLLTKMADGGGLROI	DLNFER---GGET-FLKCPVNLVVDQAPHEVDAYVEGNSKLDPTVTSFLTKMADSGGQQL	SQPAKLAIAFKYFVQGMGYMPSAMTRLMRSHTASGSSVTSILDTGRSRSHTSEGRSRSH	TQPGKLTAEAFKYFLQGMGYMASSCMTRLRSRSTABLTSAASIDGRSRSS-----	TSEGRSRSHTSEGAHLIDTRPSGAAGNSACPKNSEVYC	-----RLTSSQSSSGTLP---GGPPG---TWVESC
55	61	115	55	115	121	175	181	235	241	295	297	355	355	394	371

```

RESULT 9
SP21 HELAN
ID SP21 HELAN STANDARD; PRT; 352 AA.
AC O23969;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Pollen specific protein SP21.
GN SP21.
OS Helianthus annuus (Common sunflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; campanulids; Asterales; Asteraceae; Asteroideae;
OC Heliantheae; Helianthus.
OX NCBI_TaxID=4232;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=Flower;
RC Kruter-Caham R., Bronner R., Evrand J.L., Hahne G., Friedt W.,
RA Steimetz A.;
RT "A transmitting tissue-and pollen-expressed protein from sunflower
RT with sequence similarity to the human RTP protein.";
RL Plant Sci. 129:191-202(1997).
CC -1- TISSUE SPECIFICITY: POLLEN.
CC -1- SIMILARITY: BELONGS TO THE NRG FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y09057; CAA70260.1; -.
DR PIR; T13993; T13993.
DR InterPro; IPR004142; Ndr.
DR Pfam; PF03096; Ndr; 1.
SQ SEQUENCE 352 AA; 39191 MW; C563B2A225D4A97B CRC64;

```

Query Match	17.0%;	Score 351.5;	DB 1;	Length 352;
Best Local Similarity	29.1%;	Pred. No. 2,36-21;		
Matches	92;	Conservative	69;	Mismatches 130; Indels 25; Gaps 8;
QY	16	LVEKGETITGLQGFVDV--QEDDIETLHGSVHTLCCGPKGNRPVILTYHDIGNMKNKTCY	73	
DB	1	MADGSHISVTFPFSPFHSGGKEHIIRTICGSGSVTVCCDQE--KPDLTTPYDLALNHMSCF	58	
QY	74	NPLF--NVEDMQEIIQGHFAVCHVDAPGGQDGAASFAGVWPMRSMQQLFEMLPFGVLDQGEK	132	
DB	59	QGLFVSPESASLLAHNCFIYHINPFGHGLGASIGIDDPVPSIEDLCQILVLYNFRLG	118	
QY	133	SIIMGGTGAGAYILTRPALNNPEWVEGLVLINVPACAGNMWMAKIS-----GWTQ	185	
DB	119	SVMGCGMAAGAYILTFISIKYSEKRVTLILHSPIKCAPSWTERFYNKLTSTKLYYGWCD	178	
QY	186	ALPBMVYSHLPFKKEMQSNVEVHTTR----QHTVNDMNPENLLFLFNANYSRDLLETR	241	
DB	179	LVKSLLLHRYFSK--EVCGNPEIPESBDVYLACRKLIDERSDVNVRVYLDALDSRDDIEE-	236	
QY	242	PMFGTHVTVLTQCPALLVGDSSPAVDVVEGNSKLDPKTKLLTMADCGGLPQISQPAKL	301	
DB	237	-----LKSLECKIIIFVGDSSPHDEALQIAELKLTGNCALVEVHACGSAVVTGEQPHAM	290	
QY	302	AEAFKIVQCGM--YMP	316	
DB	291	LIPLENFLKGFGLYRP	306	

RESULT 10

YDJI_CAEEL

ID

YDJI_CAEEL

STANDARD;

PRT;

325 AA.

002485;

002485;

DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 28-SEP-2003 (Rel. 41, Last annotation update)
 DE Hypothetical 35.6 kDa protein ZK1073.1 in chromosome X.
 GN ZK1073.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Br1stol N2;
 RA McMurry A.;
 RL Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE NDRC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; 268135; CAA92227.1; -
 DR PIR; T27688; T27688.
 DR WormPep; ZK1073.1; CE03845.
 DR InterPro; IPR004142; Ndr.
 DR Pfam; PF03096; Ndr; 1.
 DR Hypothetical protein.
 SQ SEQUENCE 325 AA; 35591 MW; F68447B0A4FA74A5 CRC64;

Query Match 16.7%; Score 347; DB 1; Length 325;
 Best Local Similarity 30.4%; Pred. No. 48-21;

Matches 101; Conservative 55; Mismatches 142; Indels 34; Gaps 11;

DT 27 LQEPVQEDIEETLHGSVHTLCTGTPKGRPVILTYHDIGMNHKTCTNPLFNYEDMOEIT 86
 DB 6 LQMVVVOAONCVLHVYQGNL--BERGKTIILTVHDIGTMHKS-FRRFVNHPSMATYK 62
 QY 87 OHFAVCHVDAPGQDGAASFPAGYVYPSMDQLAEMLPVQLQFGKLSITGMGTAGAYIL 146
 DB 63 EKAIPLHVCVPGQEDNSADFFGD--FPTLDIGIDDLASVLDFEYVSATIAFEGVGANII 120
 QY 147 TRPALNPMVYGVLIINNPACBGMMDAASKI-----SGTQALPDMVSH 194
 DB 121 CRFAGHPRIWIGVILVHCTSTTLAGIIEYCKEKVNMRLSNSIMSDGA--DYLLAH 175
 QY 195 LFGKEMOSNVENVYTHROHIVNDNMPGNLHPLINAYNSRDLEIRPMPGTHVTLQCP 254
 DB 176 KFGGS-KSRQ--YBELKATINPKSLSKYLVAFTRKTLDS--STIGKLETVD-- 225
 QY 255 ALLVVGDSPPAVDAVECNKSLDPTKTTLLKMA DCGGLPOISOPAKLAFAFYVQGMGY 314
 DB 226 ALLVVGSKASHLHTVYTHKSNKKKKTLLVDNVADVQ--EAPDKLASLILCKGCGV 284
 QY 315 MFSASMTLRMSRTASGSSVLTSDGTRRSKHT 346
 DB 285 LSGVAIPGWERQRTLS-SSMEADRPBRMSVT 315

RESULT 11

BIOH_ECOLI
 ID BIOH_ECOLI STANDARD; PRT; 256 AA.
 AC P13001;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE BioH protein.
 GN BIOH OR BIOB OR B3412.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.

OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=90016899; PubMed=2678009;
 RA O'Regan M., Gloeckler R., Bernard S., Ledoux C., Ohsawa I.,
 RA Lemoine Y.;
 RT "Nucleotide sequence of the bioH gene of Escherichia coli.";
 RL Nucleic Acids Res. 17:8004-8004(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. II, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 CC -1- FUNCTION: Not clear. Seems to be implicated in the early steps of
 CC biotin biosynthesis.
 CC -----
 CC -1- PATHWAY: Biotin biosynthesis.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X15587; CAA3612.1; -
 DR EMBL; U18897; AA58210.1; -
 DR EMBL; AB000417; AAC76437.1; -
 DR PIR; Q00081; BVECBH.
 DR PDB; 1M33; 21-JAN-03.
 DR Ecogene; EG10122; bioH.
 DR InterPro; IPR000073; A/b_hydrolase.
 DR InterPro; IPR000379; Ser_gestrt_site.
 DR Pfam; PF00561; abhydrolase_1.
 DR K01 biotin biosynthesis; Complete proteome; 3D-structure.
 SQ SEQUENCE 256 AA; 28505 MW; 931226F241BBCBF3 CRC64;

Query Match 6.7%; Score 138.5; DB 1; Length 256;
 Best Local Similarity 24.2%; Pred. No. 0.0004;

Matches 69; Conservative 35; Mismatches 80; Indels 101; Gaps 14;

QY 42 GSVHTLCTGTPKGRPVILTYHDIGMNHKT--CYNPLFNYEDMOEITOHFAVCHVDAPG- 98
 DB 12 GNVHLVLT-----LHGWGLNAEVRICD-----BELSSHFTLHVLDPGF 50
 QY 99 ---QODGAASFPAGYVYPSMDQLAEMLPVQLQFGKLSITGMGTAGAYILTRPALNPE 155
 DB 51 GRSRGFGALS-----LADMEAYVLAQAPDKA-IMLGSISGLVVSQIALTLTPE 97
 QY 156 MVEGLVTLNVPACBGMMDAASKI-----ISGMTQALPDMVSHLFGKEMOSNVENVYTHYR 212
 DB 98 RVQALVTYASSPCSSARDMPGKIPDVLAGVQOQLSD-----DFQRTVE----- 141
 QY 213 OHIVNDNMPGNLHPLINAYNSRDLEIRPMPGTHVTLQCPAL--LVVGDSSPPAVDAV 270
 DB 142 -----RFLAQ--TMGTETANQDRAALKVTYALPMPREVTVL- 176
 QY 271 EONSKLDPKTTLLKMA DCGGLPOISOPAKLAFAFYVQGMGY 315
 DB 177 --NGGLEIKTVDLRQF-----LQNVSMF-----FLRYGYL 206

RESULT 12

TENA_CHICK
 ID TENA_CHICK STANDARD; PRT; 1808 AA.
 AC P10039; O73584; O73585; P11132;
 DT 01-MAR-1989 (Rel. 10, Created)


```

FT DISULFID 444 455 BY SIMILARITY.
FT DISULFID 457 466 BY SIMILARITY.
FT DISULFID 471 481 BY SIMILARITY.
FT DISULFID 475 486 BY SIMILARITY.
FT DISULFID 488 497 BY SIMILARITY.
FT DISULFID 502 512 BY SIMILARITY.
FT DISULFID 506 517 BY SIMILARITY.
FT DISULFID 519 528 BY SIMILARITY.
FT DISULFID 533 543 BY SIMILARITY.
FT DISULFID 537 548 BY SIMILARITY.
FT DISULFID 550 559 BY SIMILARITY.
FT DISULFID 564 574 BY SIMILARITY.
FT DISULFID 568 579 BY SIMILARITY.
FT DISULFID 581 590 BY SIMILARITY.
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 168 168 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 186 186 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 328 328 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 643 643 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 751 751 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 759 759 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1050 1050 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1090 1090 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1101 1101 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1112 1112 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1153 1153 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1183 1183 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1416 1416 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1736 1736 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1769 1769 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1043 1224 Missing (in isoform 2).
FT VARSPLIC 1043 1315 Missing (in isoform 3).
FT VARSPLIC 1043 1315 /ftid=VSP_001411.
FT CONFLICT 182 182 W -> R (IN REF. 2).
FT CONFLICT 571 571 SCPNDCNNV -> PAPMTATTW (IN REF. 3).
FT CONFLICT 598 598 E -> G (IN REF. 3).
FT CONFLICT 838 838 T -> TEY (IN REF. 3).
FT CONFLICT 886 886 N -> P (IN REF. 3).
SQ SEQUENCE 1808 AA; 198858 MW; B924A06CF9F5D6E CRC64;

Query Match
Best Local Similarity 5.2%; Score 107.5; DB 1; Length 1808;
Matches 75; Conservative 61; Mismatches 153; Indels 95; Gaps 18;

13 VKPLVEKGETINGLQ--EPDVQEDDIETLNG-----SVHVLCTGPKGNRPVI 59
1274 IVPGERHTHDITGLKEGEYEIE-----LYGVSSGRSROPINSVATTVGSFPKG----- 1322
60 LTYHIDIGMMHKTCTNPLFRVYEDMOETIQHFVCHYDAPGQDGAASFPAGVYRPSMDOLA 119
1333 ISFSITENSAT-----VSMTPRSRVDSYRVSYPITGCTPNVTVDGS--KTRTKLV 1374
120 EMLPGLVLOOFGKLSIIGMGTGAGAYILTRPALNPEWVEGLVLINVPCEAGMMDMAASK 179
1375 KLIVPEVDVNVNIIISVKGESESEPIGILKTALDSP---SGLVVNMITD--SEALATMOA- 1429
180 ISGWTQALPDVNVSHLFGKEKQSNVEV-----HTYRQIIVDMNPG 222
1430 ----IAAVDNYIVS--YSEDEPEVTQWVSGNTVEYDINGLARPEYTLRLVAVDAQKS 1483
223 NL--HLFINAVNSRDLIERPMGTHVTLQCPALLVVGSSPAVDAVECNS----- 274
1484 ETLSTQFTGLDAPDLATEVQSEPAVITWRP-----RAPYTDLLTYESIDGRVK 1536
275 --KLDP--TKTLLKMAOCCGLPQISQPKLAFAFK-----YFVQGMGY-MPSAS 319
1537 EYLIDPETTSYTLTELS-----PSTQYVYKLOALSRMSRSMKIQTIVFTTGLIYPPKDC 1591
320 MTRLMSRTASGSSVTSLDGTRSR 343
1592 SQALLNGEVTSGLYTIYINGDRTO 1615

```

```

RESULT 13
ID ILVC_SYNEU STANDARD; PRT; 331 AA.
AC Q8DGR0;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE ketol-acid reductoisomerase (EC 1.1.1.86) (Acetohydroxy-acid
DE isomeroeductase) (Alpha-Keto-beta-hydroxyacil reductoisomerase).
GN ILVC OR TLL2254.
OS Synecchococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synecchococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=22225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimpō S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1."
RL DNA Res. 9:123-130(2002)
CC -I- CATALYTIC ACTIVITY: (R)-2,3-dihydroxy-3-methylbutanoate + NADP(+)
CC = (S)-2-hydroxy-2-methyl-3-oxobutanoate + NADPH.
CC -I- PATHWAY: Valine and isoleucine biosynthesis; second step.
CC -I- SIMILARITY: Belongs to the ketol-acid reductoisomerase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AP005376; BAC09806.1; ALT_INIT.
CC HAMAP; MF_00435; -; 1.
CC InterPro; IPR005056; ACH_isomrctse.
CC Pfam; PF01450; ilvc; 1.
CC TIGRFAMs; TIGR00465; ilvc; 1.
CC Oxidoreductase; Branched-chain amino acid biosynthesis; NADP;
CC Complete proteome.
CC ACT SITE 108
CC SEQUENCE 331 AA; 36403 MW; E1E162D0B57D375 CRC64;

Query Match
Best Local Similarity 4.8%; Score 99.5; DB 1; Length 331;
Matches 63; Conservative 36; Mismatches 96; Indels 85; Gaps 14;

126 LQOFGKLSIIGMGTGAGAYILTRPALNPEWVEGLVLINVPCEAGMMDMAASKISGTQ 185
36 LRDSGLANVVLVYAGSSAERAR-----ABGLA---VHPVAE-----AAAGADIM 78
186 -ALPDVNVSHLFGKEKQSNVEVHTYRQIIVDMNPGHLLFINAVNSRDLIERPMP 244
79 ILDPD-----EVQRVV-----YEQETAPHLQSGNVLSFAHGFN-----IHFGQIVP 119
245 GTHT-----VTLQCPALLY-----VGDSSPAVDAVECNSK----- 275
120 PAHVADVNVADKGGHILRRTRYAQSGEVPCLFAVYQDASGARDLAMAAYAKIGSTRAGI 179
276 ----LDPYTKTLLP--KMAOCCGLPQISQPKLAFAFKFYFVQGMGYMPSASM-----TR 322
180 LETTFRETFETLDLGEQVVLCCGL-----SALIRAGFTLVQA--GYQBELAYFECIHEVK 233
323 LMSRTASGSSVTSLDGTRSRSHTSSEGRSRSHTSSEGRSR 362
234 LIVDLIVEGGIAKKRDSISNTAEVGDILTRGRIITEETRA 273

```

```

RESULT 14
MIG1_KLUMA STANDARD; PRT; 543 AA.
ID MIG1_KLUMA
AC P52288;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Regulatory protein MIG1.
GN MIG1.
OS Kluyveromyces marxianus (Yeast) (Kluyveromyces fragilis).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=4911;
RN (1)
RP SEQUENCE FROM N.A.
RA STRAIN=ATCC 12424;
RA Casart J., Ocelling J., Rome H., Delcourt J., Vandenhoute J.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INVOLVED IN GLUCOSE REPRESSION OF GLUCOSE METABOLISM
CC GENES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE CREA/MIG GROUP OF C2H2-TYPE ZINC-
CC FINGERS PROTEINS.
CC -1- SIMILARITY: Contains 2 C2H2-type zinc fingers.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; Z50018; CAA90321.1; -
DR HSSP; P08047; ISP2.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF000096; zF_C2H2; 2.
DR ProDom; PD000603; Znf_C2H2; 2.
DR SMART; SM00355; Znf_C2H2; 2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 2.
KW Transcription regulation; Repressor; DNA-binding; Nuclear protein;
KW Zinc-finger; Metal-binding; Repeat; Carbohydrate metabolism.
FT ZN FING 32 54 C2H2-TYPE 1.
FT DOMAIN 111 131 POLY-GLN.
FT DOMAIN 499 504 POLY-SER.
SQ SEQUENCE 543 AA; 59315 MW; DFC01BD92144622P CRC64;

Query Match 4.8%; Score 99.5; DB 1; Length 543;
Best Local Similarity 19.0%; Pred No. 1.7; Indels 205; Gaps 23;
Matches 98; Conservative 53; Mismatches 159;

11 AEVKEVLKGETITGLQEFVQEDITLHGSVHTLCGTPKGNRPVILTYHDIG--M 67
3 SEVVLQKKGRKAG-----AGVVVDGDEKDSRPFMCITCHRGHRL 44
68 NHKT-----CYNP--LENYEDMOEITQFAVCHVDAPG----- 98
45 EHQTSHIRTHGERPHACDFPGCAKRFGRSP--ELTRHRRIHDSKPKGKRKKKSETI 102
99 -----QDDGASFPAGWY--PSMDQLAEMLPGLVQDFGL 131
103 AREKELELQROROOQOOLQOQOQOOLQOQOHOVLPEIKISAPMASSMWEVQAINORYOD 162
132 KSIIIGMTGAGAYILTRFALNPEWEGVLVINVPACAEGMWDAASKISGMTQALPDMV 191
163 THNLSLSTVSGNSGNSGNSGNSGNSLNQVPSITSPAR-----DLA 203
192 VSHLP--GKEEMQ--SNVEVVTYRQHTVDMNPFQNLFLNA--VNSRRDLE----- 238
204 AKPMFOGSDSESESTNTTLHSVHS---QOONNGSVELLNNAARFESKSTTLNNNTP 260

```

```

RESULT 15
DHE4_CHL50 STANDARD; PRT; 523 AA.
ID DHE4_CHL50
AC P28958;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE NADP-specific glutamate dehydrogenase (EC 1.4.1.4) (NADP-GDH)
DE (Fragment).
OS Chlorella sorokiniana.
OC Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
OC Chlorellaceae; Chlorella.
OX NCBI_TaxID=3076;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=92032762; PubMed=1718478;
RA Cook J.M., Kim K.D., Miller P.W., Hutson R.G., Schmidt R.R.;
RT "A nuclear gene with many introns encoding ammonium-inducible
RT chloroplastic NADP-specific glutamate dehydrogenase(s) in Chlorella
RT sorokiniana.";
RL Plant Mol. Biol. 17:1023-1044(1991).
CC -1- CATALYTIC ACTIVITY: L-glutamate + H(2)O + NADP(+) = 2-oxoglutarate
CC + NH(3) + NADPH.
CC -1- SUBUNIT: HOMO- AND HETEROHEXAMER OF ALPHA AND BETA SUBUNITS. BOTH
CC SUBUNITS ARE ENCODED BY THE SAME GENE.
CC -1- SUBCELLULAR LOCATION: Chloroplast.
CC -1- INDUCTION: BY AMMONIUM.
CC -1- PTM: THE N-TERMINUS OF THE ALPHA AND THE BETA CHAINS ARE BLOCKED.
CC -1- SIMILARITY: BELONGS TO THE GLU/LEU/PHE/VAL DEHYDROGENASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; X58832; CAA41636.1; -
DR EMBL; X58831; CAA41635.1; ALU_SEQ.
DR FTR; S17949; S17949.
DR HSSP; P24295; LAUP.
DR InterPro; IPR006095; GLFV_dehydrog.
DR InterPro; IPR006096; GLFV_dehydrog_C.
DR InterPro; IPR006097; GLFV_dehydrog_N.
DR Pfam; PF00208; GLFV_dehydrog; 1.
DR Pfam; PF02812; GLFV_dehydrog_N; 1.
DR PRINTS; PR00082; GLFVDRGNASE.
DR PROSITE; PS00074; GLFV_DEHYDROGENASE; 1.
KW Oxidoreductase; NADP; Chloroplast.
FT NON_TER 1 1
FT ACT_SITE 202 202 BY SIMILARITY.
SQ SEQUENCE 523 AA; 57529 MW; A35FE730E5EF974 CRC64;

Query Match 4.8%; Score 99; DB 1; Length 523;

```


Best Local Similarity 21.7%; Pred. No. 1.8;
Matches 72; Conservative 50; Mismatches 136; Indels 74; Gaps 16;

```

Qy      82  MGGIIITHFAVCHVDAAGODGAASPAAGMYPSMDQLAEMPGVL-----QDFGLKSIIGM 137
Dp      225  MTELGHHISVQ--DVPAAGDIGVAAGIEIGLTFQYKRIKRYTGVLTGKGGEYGSSEIRPE 283
Qy      138  GTGAGAGVILTRPALNPF--EMVEGLVLI-----NV--NPAEGMDMAASKISGTAQLPD-- 189
Dp      284  ATGYGAVLFEVENYLKKGKGSLLKKGKCLVSGAGNAQYCAELLLEKGAIVLS-----LSDS 338
Qy      190  ---MVVSHLFGKEKEMQSNVEVVHTYRHOHIVDMNPNGNLHLEFINAANSRDLLEI--EREMPG 245
Dp      339  QGYVEBPNGETRQLQA-----YQDMKKKNNASRISEYKSDPAVYVGDRRKPM 386
Qy      246  THVTYVLCQPLLVVGGSSPAVDVAVEC--NSKLDPTKTKLLKMDCCGLLPQ--ISQPAKLA 302
Dp      387  ----ELDCC-----QVDIAFPCCATQNIIDHDALLKIKGCQYVEGAMGSPSTNE 431
Qy      303  EAEFYFVQGWGYNPSPASMTLRMSRTASGSVTSLDGTRSRSHTSBGRSRKSHTSSECTRS 362
Dp      432  AIHKYNNAGIIYCPG-----KAANAGVAVSGLEMTQNR-----MSLMTREBYRD 477
Qy      363  RSH-----TSEGAHLDTITNSCAAGNSAG 386
Dp      478  KLERIMKDIYDSAMGPSREINVDLAAGAVIAG 509

```

Search completed: January 29, 2004, 23:28:24
Job time : 19 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 23:25:38 ; Search time 41 Seconds
(without alignments)
2479.823 Million cell updates/sec

Title: US-10-057-832-2
Perfect score: 2072
Sequence: 1 MSRMQVDLAELVPLVEKQ.....TPNSGAGNSAGPKSMVESC 394

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_virus: *
16: sp_bacteriap: *
17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1646.5	79.5	323	4 Q8N959	Q8N959 homo sapien
2	1227.5	59.2	388	11 Q8VCV2	Q8VCV2 mus musculu
3	1224	59.1	363	4 Q96SM2	Q96SM2 homo sapien
4	1129	54.5	339	11 Q923D7	Q923D7 mus musculu
5	1108.5	53.5	352	11 Q8BTG7	Q8BTG7 mus musculu
6	1038	50.1	371	11 Q8V101	Q8V101 rattus norv
7	1037	50.0	357	11 Q8V100	Q8V100 rattus norv
8	1036	50.0	371	11 Q8VBW2	Q8VBW2 rattus norv
9	1035	50.0	357	11 Q8VBW2	Q8VBW2 rattus norv
10	1023.5	49.4	356	4 Q8TRD6	Q8TRD6 homo sapien
11	948.5	45.8	306	11 Q8C6E1	Q8C6E1 mus musculu
12	891.5	43.0	253	11 Q8CB00	Q8CB00 mus musculu
13	616	29.7	468	5 Q8G111	Q8G111 drosophila
14	616	29.7	485	5 Q9G5U0	Q9G5U0 drosophila
15	608.5	29.4	343	5 Q81RK7	Q81RK7 drosophila
16	608.5	29.4	447	5 Q8T0V2	Q8T0V2 drosophila

17	601.5	29.0	343	5 Q917V6	Q917V6 drosophila
18	601.5	29.0	418	5 Q81HC7	Q81HC7 drosophila
19	601.5	29.0	420	5 Q81H2	Q81H2 drosophila
20	601.5	29.0	425	5 Q917V7	Q917V7 drosophila
21	601.5	29.0	446	5 Q81HG2	Q81HG2 drosophila
22	575.5	27.8	201	11 Q8CGR9	Q8CGR9 mus musculu
23	557	26.9	109	4 Q8BR47	Q8BR47 homo sapien
24	444.5	21.5	343	5 Q9U283	Q9U283 caenorhabdi
25	388.5	18.8	320	5 Q9GU49	Q9GU49 drosophila
26	362	17.5	361	5 Q81PR2	Q81PR2 drosophila
27	362	17.5	365	5 Q9VNH7	Q9VNH7 drosophila
28	356.5	17.2	347	10 Q9ZDN1	Q9ZDN1 arabidopsis
29	355.5	17.2	364	5 Q95T61	Q95T61 drosophila
30	355.5	17.2	368	5 Q9Y164	Q9Y164 drosophila
31	339.5	16.4	344	10 Q9ASU8	Q9ASU8 arabidopsis
32	338.5	16.3	346	10 Q9FJT7	Q9FJT7 arabidopsis
33	332	16.0	356	10 Q9FU21	Q9FU21 oryza sativ
34	328.5	15.8	350	10 Q9SP24	Q9SP24 helianthus
35	312	15.1	361	10 Q9LYF2	Q9LYF2 arabidopsis
36	261	12.6	132	5 Q81H54	Q81H54 drosophila
37	230.5	11.1	201	5 Q81PR0	Q81PR0 drosophila
38	153.5	7.4	63	11 Q9CTJ5	Q9CTJ5 mus musculu
39	149	7.2	335	16 Q9HZF5	Q9HZF5 pseudomonas
40	140	6.8	315	16 Q9KJG6	Q9KJG6 pseudomonas
41	138.5	6.7	275	16 Q92AY6	Q92AY6 listeria in
42	135.5	6.5	393	16 Q9ASN6	Q9ASN6 caulobacter
43	134.5	6.5	256	16 Q8X716	Q8X716 escherichia
44	131	6.3	335	16 Q8P161	Q8P161 xanthomonas
45	128.5	6.2	275	16 Q8Y610	Q8Y610 listeria mo

ALIGNMENTS

RESULT 1
ID Q8N959 PRELIMINARY; PRT; 323 AA.
AC Q8N959;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein FLJ38330.
DE Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Kanehori K., Ishibashi T., Chiba Y., Fujimori K., Hiraoka S.,
RA Tanai H., Matanabe S., Ishida S., Ono Y., Hotura T., Matanabe M.,
RA Sugiyama T., Irie R., Otsuki T., Sato H., Ota T., Wakamatsu A.,
RA Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagaitsuma M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahara K., Masuno Y., Nagai K., Isogai T.,
RT "NEO human cDNA sequencing project";
RL Submitted (JUL-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL; AK095649; BAC04597.1; -
DR InterPro; IPR004142; Ndr.
DR Pfam; PF03096; Ndr; 2.
KW Hypothetical protein.
SQ SEQUENCE 323 AA; 35226 MW; EA130D129F3B92E8 CRC64;

Query Match 79.5%; Score 1646.5; DB 4; Length 323;
Best Local Similarity 82.0%; Pred. No. 2e-135;
Matches 323; Conservative 0; Mismatches 0; Indels 71; Gaps 1;

QY 1 MSRMQVDLAELVPLVEKGTITGLQEPVOBODITLGSVYVTLGGPKGRPYIL 60
DB 1 MSRMQVDLAELVPLVEKGTITGLQEPVOBODITLGSVYVTLGGPKGRPYIL 60
QY 61 TYHDIGMNHKTQYNPLFNVEYDMOETIOHFAVCHVDAPGQDGAASFPAQWYPSMDLAE 120

Db 61 TYHDIGNHHTC----- 72

Qy 121 MLPGVLAQFGFKLSI IGMGTGAGAYILTRPALNPEWEGVLINVPCEGMMMAASKI 180

Db 73 -----YILTRPALNNPWEVEGLVILINVPCEGMMMAASKI 109

Qy 181 SGMTQALPDWVYSHLFCKEEMOSNVEVHTYRQHIVNMNPNLHLFTNAYSRRDLEIE 240

Db 110 SGMTQALPDWVYSHLFCKEEMOSNVEVHTYRQHIVNMNPNLHLFTNAYSRRDLEIE 169

Qy 241 RPMGHTTVILQCPALLVVGDSSPAVDAVECNKSLDPTKTLIKMADCGGLPQISOPAK 300

Db 170 RPMGHTTVILQCPALLVVGDSSPAVDAVECNKSLDPTKTLIKMADCGGLPQISOPAK 229

Qy 301 LAEFKFFVQGMGYMPSASWTRLMRSLTASGSSVTSLDGTRSRHTSGTSSRHTSEGT 360

Db 230 LAEFKFFVQGMGYMPSASWTRLMRSLTASGSSVTSLDGTRSRHTSGTSSRHTSEGT 289

Qy 361 RSRHTSEGAHLDTITNSGAGNAGSAPKAMEVSC 394

Db 290 RSRHTSEGAHLDTITNSGAGNAGSAPKAMEVSC 323

RESULT 2	OBVCV2	PRELIMINARY;	PRT;	388 AA.
ID	OBVCV2			
AC	OBVCV2;			
DT	01-MAR-2002 (TREMBlrel. 20, Created)			
DT	01-MAR-2002 (TREMBlrel. 20, last sequence update)			
DT	01-OCT-2002 (TREMBlrel. 22, last annotation update)			
DE	Similar to N-myc downstream regulated 3.			
GN	NDR3.			
OS	Mus musculus (mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxId=10090;			
RN	[1]			
RC	SEQUENCE FROM N.A.			
RC	TISSUE=Retina;			
RA	Strausberg R.;			
RL	Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; BC018504; AAH18504.1; -			
DR	MED; MGI:1352499; Ndr3.			
DR	InterPro; IPR004142; Ndr.			
DR	InterPro; IPR000379; Ser_estrs_site.			
DR	Plam; PF03096; Ndr; 1.			
SQ	SEQUENCE 388 AA; 42982 MW; AD64FE1A627469C0 CRC64;			
Query March	59.2%;	Score 1227.5;	DB 11;	Length 388;
Best Local Similarity	59.1%;	Pred No. 1,1e-98;		
Matches 241;	Conservative 58;	Mismatches 70;	Indels 39;	Gaps 7
QY	4 EMQDVDLAEVPLVEKGETITGLLOEFDVQEDIDETLHGSVHYTLCGTPKGNRPVITLTH 63			
DB	3 ELQDVQVLEIKPIL-NDKNGTRNFQDFDCQEHDIETPRGHWVITIRGLPKGNRPVITLTH 61			
QY	64 DIGNNHTCTCNPLFENVEDMOETQHFHGVCHVAPQOQOAGASFPAGYNYPEMDLAELP 123			
DB	62 DIGNNHSCNTEFNEDMOETQHFHGVCHVAPQOQOAAFSFPGYQYPTMDLAELMP 121			
QY	124 GVLQOQFGKSLIOMGAGAGAYILTRPALNPPMEVGLVILVNPFCAGGMDMAASKISGW 183			
DB	122 PVLTHLSMSKSLIGIGAGAGAYILSRPALNHPRLVGLVILNIDPCAKGMDMAASKISGF 181			
QY	184 TQALPDMVNVSLFGKEKQMSNVEVHTYRQHIIVNDMPGNLHLPLINAYNSRRDLEIERP 243			
DB	182 TTNIVDIILAHFGQSELANDLIGTYRLHIAQINDENILQFLGYSNGRRDLEIERPI 241			
QY	244 PG---THYVTLQCPALLVGDSSPANDVAVEGNSLTDPTKTLTKMADCGGLPDISPAK 300			
DB	242 LGQNDRLKTLKCKSTLLVGVNDGNSPAVEAVECNSRLDPIINTLLKMDCCGLPDVVQPGK 301			

```

Oy 301 IAAAFYCFPOAKGY-----MPSASMTRLKRSRT-AGGSVTLSDGRSSHT 346
Db 302 LTAAFKYFLQAGMOYIPFYQLSHLSSSESPASASMTRLKRSRHTSSSSIGGGEPSRSTV 361
Oy 347 SEGTNRSRHTSEGTNRSRHTSGAHLDTTPNSGAAGSACKMEVRC 394
Db 362 -----SNQSDGTQSCSP-----DVLDRH-----QTLMEVRC 388

```

RESULT 3	Q96SM2	PRELIMINARY;	PRT;	363 AA.
AC	Q96SM2;			
DT	01-DEC-2001 (TrEMBLrel. 19, Created)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)			
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)			
DE	Hypohectical protein FLN14759.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX	NCBI_TaxId=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Isoegl T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,			
RA	Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,			
RA	Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,			
RA	Arata M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,			
RA	Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y., Oshima A.,			
RT	"NEDD human cDNA sequencing project."			
RL	Submitted (May-2001) to the EMBL/Genbank/DBJ databases.			
DR	EMBL; AK027665; BAB55227.1; -			
DR	InterPro; IPR004142; Ndr.			
DR	InterPro; IPR000379; Set_estrs_site.			
DR	Pfam; PF03096; Ndr; 1.			
DR	Hypohectical protein.			
QO	SEQUENCE 363 AA; 39922 MW; BA1B490BE1AE79C1 CRC64;			

```

Query Match 59.1% Score 1224; DB 4; Length 363;
Best Local Similarity 60.5%; Pred. No. 2e-98;
Matches 239; Conservative 56; Mismatches 62; Indels 38; Gaps 6

QY 4 EMQVDLAEEVKEFLVEKGERTITGLLOEERPDVQODIETLHGSHVHTLCCGTPKGNRPVILTYH 63
Db 3 ELQDVQVLEIKRPLND-----KEHDIETHGVHVHTIRGLPKGNRPVILTYH 49

QY 64 DIGNHNKTCINPLEFNVEDMOEITQHPAVCHYADAGQODGAASFPAGTMYRPSMDLAEMLP 123
Db 50 DIGNHNKSCSAAPFNPFDMOEITQHPAVCHYADAGQODGAASFPAGTQYPTMDLAEMLP 109

QY 124 GVLDQFGLKSIIGKGTAGAYILTRFPLNNEBMEGVLIVNNECAEEMDMAASKISGM 183
Db 110 PVLTHLSKSIIGTGVAGAYILSRFALNHVELVEGLVILNVDCAKGIDMAASKISGL 169

QY 184 TQALPDMVNSHFGKEEMQSNVEVHTYRQHIYNDMNPGLHFLINAYNSRDLIERPM 243
Db 170 TTVNVVDIILAHNFQGEELQANLDDIQTRMHADIQNDNQLEFINSYNGRRDLIERPI 229

QY 244 PGTH---TVLQCCALLVQDSSPAVDAYVECNKLPPTKTTLLKMDCCGLPOISOPAK 300
Db 230 LGQDNKSKTLKSCSTLLVQDSSPAVDAYVECNKLPINITLLKMDCCGLPOVQCGK 289

QY 301 LAEAFKYFVQGMGYMPASMTRLMRKRT-ASGSSVTSIDGTRSRSHSTSEGTRSRSHSTSEG 359
Db 290 LTFEAFKFLQGMGYIPASMTRLMRKSTHSTSSLSGGESEFSSVTSV-----SNGSDG 342

QY 360 TRSRSHSTSEGALHDIITPNSGAAGNSACKPKMEVSC 394
Db 343 TQESCESP-----DVLDRH-----QTMVEVSC 363

RESULT 4
ID 0923D7 PRELIMINARY; PRT; 339 AA.

```

AC Q933D7, 2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
 DE Similar to NDRG family, member 4.
 GN NDR4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RX [1]
 RN
 RP
 RP SEQUENCE FROM N.A.
 RA Straubeberg R.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC006595; AA06595.1; -
 DR MGD; MGI:2384590; Ndr4.
 DR InterPro: IPR004142; Ndr.
 DR Pfam; PF03096; Ndr; 1
 SQ
 SQ SEQUENCE 339 AA; 37165 MW; 4FE9617B206700A2 CRC64;

Query Match	54.5%;	Score 1129;	DB 11;	Length 339;
Best Local Similarity	63.5%;	Pred. NO. 3.5e-90;		
Matches 209; Conservative	50;	Mismatches 66;	Indels 4;	Gaps 1;

Qy	34	EQDIEIETHGSHVHTLCGTGPKGRNPVLLIYTHDIGNHNTCSYBPLFYEDMOBITOHFAVCH	93
Db	8	EHDIETFPYGLHNVLYINGSPKGRNPAILTYHDVGLGNHKSICFTPFNFEDMOBITGHPVCH	67
Qy	94	VDAFGOODGAASFPAGTMYTFSMDOLAEMLPGYLOOFGELKSIIHGCTGAGATLTRPILANN	153
Db	68	VDAFGOOVAGSOPFOOYOPFSPMEOLANMLPSVVOHFGRKYVIGIGVGAGAVLAKFALIF	127
Qy	154	PEWEGIGVLINVPSCAEGMDMAASKISGMTOLBDMVSHLFGKEKQOSNVEVHTYRQ	213
Db	128	PDLVBGLVLMNIDPNKGMIIDWAAITLGSLTJSTLPDYLSHLSFQEBELVNNTELVSQYRQ	167
Qy	214	HYNDMHPGULHFTIYAAYNSRRDLBELERMPGTHVTYIQCPRLLVYGGSSPAVDAYVECN	273
Db	188	QISNVNQALQLFMMYNSRRDLDIRNGVTPNAKTLRCVPMILVGGNNAAEAGVECN	247
Qy	274	SKLDPTKTLTLKKAADCGGLPOISOPAKLAEAKEYVVOGMYWPSASMTLRMRSTAGSS	333
Db	248	SKLDPTTTTLTKKADSGGLPYVTQPKELTEAKYFTLQGGMYWPSASMTLRMRSTATLTS	307
Qy	334	VTSLDGR-----SRSHTSRGTSRSRSHTS	358
Db	308	ASSVDSGRPQCAHSDSSSGMGQVNNHME	336

RESULT 5

ID	OBERTG7	PRELIMINARY;	PRT;	352 AA.
AC	OBERTG7;			
DT	01-MAR-2003 (TrEMBLrel. 23, Created)			
DT	01-MAR-2003 (TrEMBLrel. 23, last sequence update)			
DT	01-MAR-2003 (TrEMBLrel. 23, last annotation update)			
DE	NRG4 protein.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euteleota; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Medulla oblongata;			
RX	MEDLINE=22354683; PubMed=12466851;			
RA	The PANTOM Consortium,			
RA	the RIKEN Genome Exploration Research Group Phase I & II Team;			
RT	"Analysis of the mouse transcriptome based on functional annotation of			
RL	60,770 full-length cDNAs."			
RL	Nature 420:563-573(2002).			
RQ	EMBL; AK090374; BAC1189.1; -			
SR	SEQUENCE 352 AA; 38509 MW; 3ED388148DC852CD CRC64;			

Query Match	53.5%	Score 1108.5;	DB 11;	Length 352;
Best Local Similarity	60.8%;	Pred. No. 2.3e-88;		
Matches 208; Conservative	51;	Mismatches 66;	Indels 17;	Gaps 2;

QY	34	BODIETLHSGSVHTLQCPKRGKRPVILTYHHIGMNHKTCVMPYLPNYSDEMOEITQHFVCH	93
Db	8	EHDIETPPYGLLHVILRGSGPKNRPAILTYHHVGANLHKCFETFPNPFEDMOEITGHFVCH	67
QY	94	VDAPGQDDAASFPFAGVWYPSMDQJLAEMLPVLOQFGLKSITIGMGTAGAVIILTRFALNN	153
Db	68	VDAPGQGVASQFPQGYQFPSMEQLAANLPBVQHHFGKRYIIGVGAGAVLAKPILIF	127
QY	154	PEWVEGLVILINVPCEAGMMDWAASKISGWTQALPDWVSHLFGKEEMQSNVEVHTYRQ	213
Db	128	PDVEGLVAMNIDPBGKGMIDWAATKSLGSLTLPDVLSHLFSQEBLVNNTVELVSYRQ	187
QY	214	HYVNDMPENLHLFINAVNSRRDIEIERPMGCHTVTLQCPALLVGDSSPAVDAYVECN	273
Db	188	QISNVNQANLQLFNMMYNSRRDLDINRPGVPAKTLRCFVMLVVDGNAPAEGVVECN	247
QY	274	SKLDPKHTLLKMAOOGGLPOLSOPAKLAAPKFFVQMGV-----MPGASM	320
Db	248	SKLDPTTITFLMAASGGLPVQYQGLKTEAFKFLQOMGIAHLKDRRLSGGAVPSASM	307
QY	321	TRLMRSTASGSSVTSLOSTR-----SRSTSEGRSRSHTE	358
Db	308	TRLMRSTASVTSASVDSGRPOPCAHSDSDSGMGQVNHTE	349

RESULT 6	
Q8VI01	
ID Q8VI01	PRELIMINARY; PRT; 371 AA

DT 01-MAR-2002 (TEMBLrel . 20, Created)
DT 01-MAR-2002 (TEMBLrel . 20, last sequence update)
DT 01-JUN-2002 (TEMBLrel . 21, last annotation update)
DE Antidepressant-related protein ADRG123.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Takahashi K., Yamada M., Yamada M.;
RT "Identification of a novel antidepressant related gene, ADRG123, after
RT chronic antidepressant treatment in rat frontal cortex";
RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL, AF341405, AAL73186.1, --
DR InterPro, IPR004142, Ndr.
DR Pfam, PF03096, Ndr; 1.
SQ SEQUENCE 371 AA; 40764 MW; C049A683F37D1DAA CRC64;

Query Match	50.1%;	Score 1038;	DB 11;	Length 371;
Best Local Similarity	51.6%;	Pred. No. 13.6e-82;		
Matches 206; Conservative	65;	Mismatches 90;	Indels 36;	Gaps 6;

QY 4 EMDVDVLAIEVKPLV-----EKGEFTTGLLOEDVDVQEDDIEFTLHSSVHYTLQCTPKGN 55
 Db 3 ELIOEVITETKPLLPQGPPEAKAEALARIILLD-QGQTHSVEKPRYSVFTVYGTGPK 61
 QY 56 RPVILTYHDIIGNMKTCVYNPLFNYEDMOEITOHFAVCVADAPGQDQGAASEFPAQWYPSM 115
 Db 62 RPAIFTYHVGUNYKSCFQPLFPQFDOMEIIONFRAHYVADAPGMBEQAPVPLGYQPSL 122
 QY 116 DQIAEMLPGVLOQFGKSLITGMGTGAGAYILTRPALNPEKVBGLVILNPPCABEGMDW 175
 Db 122 DQIADMDIPCLLYLNFSTIIIGVGAGAYILSRVALNPDVTEGVLVINDPNKGMMDW 181
 QY 176 AASKIGMTQALPDVNVSHLPGEKEMQSNVEVPHYROIYNDQNPQGNLHFIAYNYSRR 235
 Db 182 AAHKLTGLTSLSPENILGHLFQSEELSGNSEILQKTRSLITHARLENLIELYNNYSYNNRR 241
 QY 236 DLEIERPMEGTHVTYLQCPALLVAVGDSSPAVDAVECNKSLDPTKTYLLKMAQCGSLPQI 239

Db 242 DLFNR-----GSEMTLKCPVMLVVGDOA.PHEDAVVECNKSLDPTQTSFLKMAADSGGQQL 297
 Qy 296 SOPALAAAFKFFVQGMGMYMSASMTLRMRSTRASGSSVTSLDGTRSRSHTSEGRSRSH 355
 Db 298 TOPGLTAAFAKFFVQGMGMYMSASMTLRMRSTRASGSSVTSLSAASIDGSRSS----- 346
 Qy 356 TSEGRSRSHSEGAHLDTTPNSGAAGNSAGPKSMEVSC 394
 Db 347 -----RTLSQSSSESGTLP-----SGPPGH-----TMEVSC 371

RESULT 7

08V100 PRELIMINARY; PRT; 357 AA.
 ID 08V100
 AC 08V100.
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Antidepressant-related protein ADRG123 splice variant.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Takahashi K., Yamada M., Momose K.;
 RT "Identification of a novel antidepressant related gene, ADRG123, after
 chronic antidepressant treatment in rat frontal cortex."
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF334106; AAL73187.1; -
 DR InterPro; IPR004142; Ndr.
 DR Pfam; PF03096; Ndr; 1.
 SQ SEQUENCE 357 AA; 39255 MW; 66792ABEA9DD9875 CRC64;

Query Match 50.0%; Score 1037; DB 11; Length 357;
 Best Local Similarity 52.4%; Pred. No. 4.1e-82;
 Matches 205; Conservative 63; Mismatches 87; Indels 36; Gaps 6;

Qy 4 EMQDVDAEVRPLVKEGTTIGLQEPVQEDITLHGSVHTLCGTPKGNRYLITLH 63
 Db 3 ELQEVQITEKRPPLPGQTPPEA-----AKTHSVETPPGVSFTVYGTTPKRPRAIFTH 55
 Qy 64 DIGMNHKTCYNPLFVNEBMOEITQHFVACHVADAPQODGAASPPAGYWPSPMDLAEMLP 123
 Db 56 DVLGNYKSCFQPLPFGDMQELIQFVAVHVDAPGMEGAPVFPGLGYPSLDQADMIIP 115
 Qy 124 GVLQOFGKLSIIGMGTGAGAYILTRFALNPEMVEGLVINVPACBGMMDMAASKISGW 183
 Db 116 CILQYLNSTIIIGVAGAYILSRVYALNHPDVEGLVLINIDPAKGMMDMAAKLTLGL 175
 Qy 184 TQALPDMVSHLFGKEBQMSQVNEVHTTRQHI VDNMNGNHLFLINAVNSRDEIEFRPM 243
 Db 176 TTSIPEMILGHLFSGEELSGNSELIQKXRSITTHAPNLENTIELYNSNNNRDINFR-- 233
 Qy 244 PGTHTVTLQCPALVVGDSPAVDAVVECNKSLDPTKTLTKMAADCGGLPOISQPAKLA 303
 Db 234 --GSEMTLKCPVMLVVGDOA.PHEDAVVECNKSLDPTQTSFLKMAADSGGQQLTOPGLTTE 291
 Qy 304 AFKFFVQGMGMYMSASMTLRMRSTRASGSSVTSLDGTRSRSHTSEGRSRSH 363
 Db 292 AFKFFVQGMGMYMSASMTLRMRSTRASGSSVTSLSAASIDGSRSS-----RTL 335
 Qy 364 SHTSEGAHLDTTPNSGAAGNSAGPKSMEVSC 394
 Db 336 SQSSSESGTLP-----SGPPGH-----TMEVSC 357

RESULT 8

08VBW2 PRELIMINARY; PRT; 371 AA.
 ID 08VBW2
 AC 08VBW2.
 DT 01-MAR-2002 (TREMBlrel. 20, Created)

DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE NDRG1 related protein NDRG2b1 (NDRG1 related protein NDRG2a1).
 GN NDRG2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Heart;
 RA Boulkroun S., Fay M., Zennaro M.C., Escoubet B., Jaisser F.,
 RA Blot-Chabaud M., Farman N., Courtois-Couty N.;
 RT "Identification of rat N-myc Downstream-Regulated Gene 2 as a novel
 early aldosterone-induced gene."
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ246426; CAD19999.1; -
 DR EMBL; AJ246424; CAD19997.1; -
 DR InterPro; IPR004142; Ndr.
 DR Pfam; PF03096; Ndr; 1.
 SQ SEQUENCE 371 AA; 40779 MW; C049A1AAAF37D115A CRC64;

Query Match 50.0%; Score 1036; DB 11; Length 371;
 Best Local Similarity 51.6%; Pred. No. 5.4e-82;
 Matches 206; Conservative 64; Mismatches 91; Indels 38; Gaps 6;

Qy 4 EMQDVDAEVRPLV-----EKGTITGLQEPVQEDITLHGSVHTLCGTPKGN 55
 Db 3 ELQEVQITEKRPPLPGQTPPEAAEALAAARILD-QQTHSVETPPGVSFTVYGTTPKPK 61
 Qy 56 RPVLITVTHDIGMNHKTCYNPLFVNEBMOEITQHFVACHVADAPQODGAASPPAGYWPSPM 115
 Db 62 RPAIFTHDVLGNYKSCFQPLPFGDMQELIQFVAVHVDAPGMEGAPVFPGLGYPSQ 121
 Qy 116 DQLAEMLPVGLQOFGKLSIIGMGTGAGAYILTRFALNPEMVEGLVINVPACBGMMDW 175
 Db 122 DQADMIIPCLIQYLNSTIIIGVAGAYILSRVYALNHPDVEGLVLINIDPAKGMMDW 181
 Qy 176 AASKISGWTQALPDMVSHLFGKEBQMSQVNEVHTTRQHI VDNMNGNHLFLINAVNSRR 235
 Db 182 AAHLTGLTSSIPEMILGHLFSGEELSGNSELIQKXRSITTHAPNLENTIELYNSNNNR 241
 Qy 236 DLBIETPMPGTHTVTLQCPALVVGDSPAVDAVVECNKSLDPTKTLTKMAADCGGLPOI 295
 Db 242 DLFNR-----GSEMTLKCPVMLVVGDOA.PHEDAVVECNKSLDPTQTSFLKMAADSGGQQL 297
 Qy 296 SOPALAAAFKFFVQGMGMYMSASMTLRMRSTRASGSSVTSLDGTRSRSHTSEGRSRSH 355
 Db 298 TOPGLTAAFAKFFVQGMGMYMSASMTLRMRSTRASGSSVTSLSAASIDGSRSS----- 346
 Qy 356 TSEGRSRSHSEGAHLDTTPNSGAAGNSAGPKSMEVSC 394
 Db 347 -----RTLSQSSSESGTLP-----SGPPGH-----TMEVSC 371

RESULT 9

08VBW2 PRELIMINARY; PRT; 357 AA.
 ID 08VBW2
 AC 08VBW2.
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE NDRG1 related protein NDRG2b2 (NDRG1 related protein NDRG2a2).
 GN NDRG2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Heart;
 RA Boulkroun S., Fay M., Zennaro M.C., Escoubet B., Jaisser F.,
 RA Blot-Chabaud M., Farman N., Courtois-Couty N.;

"Identification of rat N-myc Downstream-Regulated Gene 2 as a novel
RT easily aldosterone-induced gene."

RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AJ426427; CAD0000.1; -
DR EMBL: AJ426425; CAD0000.1; -
DR InterPro: IPR004142; Ndr.
DR Pfam: PF03096; Ndr; 1.
SQ SEQUENCE 357 AA; 39270 MW; 66792DC7A9DD9485 CRC64;

Query Match 50.0%; Score 1035; DB 11; Length 357;
Best Local Similarity 52.4%; Pred. No. 6.2e-82;
Matches 205; Conservative 62; Mismatches 88; Indels 36; Gaps 6;

QY 4 EMQVDLAEPVLEKEGTTGLLQEPVQODIETTLGSHVTLCTPKNRPVILTYH 63
DB 3 ELQVQITREKPL-PGOTPEA-----AKTHSVETPGSVTFYVYGPCKRPAIFTYH 55
QY 64 DIGNNHKTCTNPLFNYEDMOEITQHFVAVCHVDAPQODGAASFPAGYMPMDQLAELP 123
DB 56 DVLGNYKSCFQPLFQFGDMQEIIGFVAVHYDAPGMEGAVPFLGYQPSQDQLADWIP 115
QY 124 GVLDQFGKSLTIGMGTGAGATLFRFALNPEMVEGLVINVNPAEGMMMAASKISGM 183
DB 116 CILDYLNFTIIGVGAGAYILSRVALNHPDVEGLVINIDPAKGMMDMAHKLTL 175
QY 184 TQALPDVNVSHLFGKEEQSNVEVHTYRQHVNDMPGNLHLPINAVNSRRDLEIERPM 243
DB 176 TSSIPEMILGHLFSGEELSGNSELIQKTRSLTHAPULENIELYNSTNNRDLNFER-- 233
QY 244 PGTHTVTLQCPALLVVDSSPAVDVAVECNSKLDPTKTLTKMDCGLPOISQPAKLA 303
DB 234 --GGMTLKCPVMLVVGQAHPEDAVVECNKLDPTQSFVKMADSGQPOLQPGKLT 291
QY 304 AFKTFVQMGVMPASMTRLMRSTASGVTSLSLDTGTSRSHTSRSTSGTSRSTSGTSR 363
DB 292 AFKTFVQMGVMASSCMTRLSRSTASGVTSLSLDTGTSRSHTSRSTSGTSRSTSGTSR 335
QY 364 SHTSEGAHLDTIPNSGAAGNSAGPKSMEVSC 394
DB 336 SSSSESGTLP-----SGPPGH-----TMEVSC 357

RESULT 10

Q8TRD6 PRELIMINARY; PRT; 356 AA.

AC Q8TRD6 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Cytoplasmic protein Ndr1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhou Y., Yu L., Zhao Y., Fan Y.X., Huang J., Zhao S.Y.;
RT "Cloning and expression of a novel human cDNA homology to murine
cytoplasmic protein Ndr1 mRNA."
RL Submitted (Apr-2002) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Yue P., Yu L., Ding J.B., Zhao Y., Huang J., Zhao S.Y.;
RL Submitted (Aug-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF087872; AAM10500.1; -
DR InterPro: IPR004142; Ndr.
DR Pfam: PF03096; Ndr; 1.
SQ SEQUENCE 356 AA; 39531 MW; E9470B6A701B9C17 CRC64;

Query Match 49.4%; Score 1023.5; DB 4; Length 356;
Best Local Similarity 55.3%; Pred. No. 6.2e-81;
Matches 198; Conservative 56; Mismatches 75; Indels 29; Gaps 4;

QY 37 IETLHGSVHVTLCSTPKGNRPVILTYHDIGNNHKTCTNPLFNYEDMOEITQHFVAVCHVDA 96
DB 28 VETFGSVTFVYGTGPKRPAITFDVGLNYSQFPLFQFGDMQEIIGFVAVHYDAPGMEGAVPFLGYQPSDQLAD 87
QY 97 PGQODGAASFPAGYMPMDQLAELPFGVLDQFGLKSLTIGMGTGAGATLFRFALNPEM 156
DB 88 PGMEGAPVPELGYQPSDQLADIMPCVLOLYNFTSIIGVGAGAYILSRVALNHPD 147
QY 157 VEGVLINVNPAEGMMMAASKISGWTQALPDVNVSHLFGKEEQSNVEVHTYRQHV 216
DB 148 VEGVLINVNPAEGMMMAASKISGWTQALPDVNVSHLFGKEEQSNVEVHTYRQHV 207
QY 217 NDMPGNLHLPINAVNSRRDLEIERPMPGTHTVTLQCPALLVVDSSPAVDVAVECNSKL 276
DB 208 HAPLNDIELYNSTNNRDLNFER-----GGDITLRCPVMLVVGQAHPEDAVVECNKLD 263
QY 277 DPTKTLTKMDCGLPOISQPAKLAELPFGVQMGVMPASMTRLMRSTASGVTS 336
DB 264 DPTQTSFLKMAADSGQPOLQPGKLTFAFKYFLQMGVMASSCMTRLSRSTASGVTS 323
QY 337 LDGTRSRSHTSRSTSGTSRSTSGTSRSTSGTSRSTSGTSRSTSGTSRSTSGTSRSTSGTSR 394
DB 324 VDGNRSHS-----RTLSQSSSESG-----TSSGPPGH-----TMEVSC 356

RESULT 11

Q8C661 PRELIMINARY; PRT; 306 AA.

AC Q8C661 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE N-myc downstream regulated 2 (Fragment).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL: AK076514; BAC36373.1; -
DR NON TER 1
FT 1
SQ SEQUENCE 306 AA; 33701 MW; 42B01C152E1972F CRC64;

Query Match 45.8%; Score 948.5; DB 11; Length 306;
Best Local Similarity 55.4%; Pred. No. 1.8e-74;
Matches 185; Conservative 53; Mismatches 67; Indels 29; Gaps 5;

QY 61 TYHDIGNNHKTCTNPLFNYEDMOEITQHFVAVCHVDAPQODGAASFPAGYMPMDQLAE 120
DB 2 TYHDVGLNYSQFPLFQFGDMQEIIGFVAVHYDAPGMEGAVPFLGYQPSDQLAD 61
QY 121 MLPEVLQFGLKSLTIGMGTGAGATLFRFALNPEMVEGLVINVNPAEGMMMAASKI 180
DB 62 MIPCLQYLNFTIIGVGAGAYILSRVALNHPDVEGLVINIDPAKGMMDMAHKL 121
QY 181 SGWTQALPDVNVSHLFGKEEQSNVEVHTYRQHVNDMPGNLHLPINAVNSRRDLEIE 240
DB 122 TGLTSSIPDMILGHLFSGEELSGNSELIQKTRGILTHAPULENIELYNSTNNRDLNFB 181
QY 241 RMPGTHTVTLQCPALLVVDSSPAVDVAVECNSKLDPTKTLTKMDCGLPOISQPAK 300
DB 182 R---GGET-TLKCPVMLVVGQAHPEDAVVECNKLDPTQSFVKMADSGQPOLQPGK 237
QY 301 LAEPKTFVQMGVMPASMTRLMRSTASGVTSLSLDTGTSRSHTSRSTSGTSRSTSGTSR 360
DB 238 LTEAFKTFVQMGVMASSCMTRLSRSTASGVTSLSLDTGTSRSHTSRSTSGTSRSTSGTSR 281

QY 361 RSRSHSEGALDITPNSGAGNAGAPKSMVEVSC 394
 DB 282 RLSQSSSESGLT-----SGPCH-----TMEVSC 306

RESULT 12

Q8CBDO PRELIMINARY; PRT; 253 AA.

QY 08CBDO: 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE N-myc downstream regulated 3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RP SEQUENCE FROM N.A.
 RX STRAIN=C57BL/6J; Tissue=Cerebellum;
 RC MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL: AK036227; BAC29369.1; -.
 SQ SEQUENCE 253 AA; 28583 MW; 568DCTF547BDF0A3 CRC64;

Query Match 43.0%; Score 891.5; DB 11; Length 253;
 Best Local Similarity 66.1%; Pred. No. 1.3e-69;
 Matches 160; Conservative 40; Mismatches 41; Indels 1; Gaps 1;

QY 4 EMQVDLAEVKPLEKGETITGLLOEPVODIETLHGSVHTLCGPKGNRPVILTYH 63
 DB 3 ELQDVLLEIKRPL-NDKNGTRNFQDFCOEHDIETPHGMVHTIRGLPKGNRPVILTYH 61
 QY 64 DIGMNHKTCYNPLPYEDMOEITQHFVAVCHVDAPQODGAASFPAGWYPSMDLAEMLP 123
 DB 62 DIGMNHKTCYNPLPYEDMOEITQHFVAVCHVDAPQODGAASFPAGWYPSMDLAEMLP 121
 QY 124 GVLQDFGLKSIIGMGTGAGAYILTRFALNPEMVEGLVINVPACBGMMDMAASKISGW 183
 DB 122 PVLFTLSMKSIIGIGVAGAYILSRFALNPELVEGLVINIDPCAKIMDMAASKISGF 181
 QY 184 TQALPDWVSHLFGKEEMOSNVEVHTYRQHIIVNDMPGNLHLFINAVNSRDLIERPM 243
 DB 182 TTNIYVDIILAHFGQELQAMLDIOTYRLHIAQDINOENLQFLGYSNGRRDLIERPI 241
 QY 244 PG 245
 DB 242 LG 243

RESULT 13

Q8G11 PRELIMINARY; PRT; 468 AA.

QY 08G11: 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE RH13074P.
 OS Mesk2.
 OC Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RP SEQUENCE FROM N.A.
 RC STRAIN=Y;
 RA Stapleton M., Brokstein P., Hong L., Agdayani A., Carlson J.,

RA Chame M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
 RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
 RA Ceiniker S.,
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BT001773; AAN71528.1; -.
 SQ SEQUENCE 468 AA; 51588 MW; E3CDA148332DDCC3 CRC64;

Query Match 29.7%; Score 616; DB 5; Length 468;
 Best Local Similarity 35.0%; Pred. No. 3.8e-45;
 Matches 151; Conservative 62; Mismatches 162; Indels 56; Gaps 11;

QY 5 MODVDLAEVKPLEKGETITGLLOEPVODIETLHGSVHTLCGPKGNRPVILTYH 64
 DB 6 MDDELASVGLQFPNAR---GSILF-ACEGRRPVTDGVDHVAIQGTA--KRALITYHD 59
 QY 65 IGMMHKTCTNPLPYEDMOEITQHFVAVCHVDAPQODGAASFPAGWYPSMDLAEMLP 124
 DB 60 LGINVATSFAGFFNFPVWRGLLENFCVYHTARQGEAGPTLPEDVYPTMDLAAQLLF 119
 QY 125 VLOQDFGLKSIIGMGTGAGAYILTRFALNPEMVEGLVINVPACBGMMDMAAS----- 178
 DB 120 VLSHFGKSVIIGGVGAGAYILTRFALNPEMVEGLVINVPACBGMMDMAAS----- 179
 QY 179 -KISGWTQALPDWVSHLFGKEEMOSNVEVHTYRQHIIVNDMPGNLHLFINAVNSRDL 237
 DB 180 LRTKGMVQSVITDILMHFGRNBERNHDVQWTKQHFEGVAPFTLAILNIYIRNDL 239
 QY 238 EIERPMFG-----THVTTLQCPALLVGDSSPAVDVAVECNSKLDPTKTLKMAADCG 292
 DB 240 HLRATPPTGTSERATATLTKMPVINITGSLSPHYDDVTFTNGRDLPTNSMMKISDC-AL 298
 QY 293 PQISQPKLAELAKYFPVQNGY-----MPSAS-----MTRL 323
 DB 299 VLEBQPKLAELARFLFLOGGGYATPLSTPASSCGKTYHTYSSIFANFREDOQAAMEER 358
 QY 324 MRSRTASGSSVTSID---GTRSRSHTEGTRSRSHTEGTRSRSHTE---EGALHDITP 376
 DB 359 DRERERDRDQROLSLVRGRRLRRLTAINCTSSNNTTGAAGGDNNMADRSDVDDQDLN 418
 QY 377 NSG-AAGNSAG 386
 DB 419 GNGCAGGNGTG 429

RESULT 14

Q9GU50 PRELIMINARY; PRT; 485 AA.

QY 09GU50: 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Msexpression suppressor of KSR (GH09802p).
 GN Mesk2 OR CG15668 OR CG15669.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 RP SEQUENCE FROM N.A.
 RC STRAIN=Y;
 RA Stapleton M., Brokstein P., Hong L., Agdayani A., Carlson J.,
 RA Chame M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
 RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,


```

Db      1 MDIEILRSVQLQFPNAR---GSILE-ACEQRVRVPTDKGDVHVAIQGDTA--KPAIITYHD 54
QY      65 IGMNHTCYNPLFNYEDMOEITTOHFAVCHVDAFGQDGAASFPAQYMYPSMDLAEMLP 124
Db      55 LGLNVAISFAGFENFPVVRGLENFCYHVTAPGQEBGAPTLPEYVYPTMDDLAQULF 114
QY      125 VLOQFGLKSIIGMGTGAGAYILTRFALNPPMEVEGLVLINVPCEAGMDWAAS----- 178
Db      115 VLSHFGKSVIGFVGAGANILARFAHAHPDKVGCALCLINCSTQSGWIEKGYQSFNARF 174
QY      179 -KISGWTQALPDVNVVSHLFGKEEMQSNVEVVTYRQHI VDNMPGNLHLFTNAYNSRRDL 237
Db      175 LRTGWTQGVIDYLMHMFGRNPEERHNDLVQMTYKQHFBERGVNPTNLAMLINAYIHRNDL 234
QY      238 EIERPMPG-----THVTLOCPALLVGDSSPAVDAVECNSKLDPTKTLTKMADCGL 292
Db      235 HLAHTPPGTGSETAATTLKMPVINITGSLSPHYDDTVTFNGRLDPTNSSWMKISDC-AL 293
QY      293 POISQPAKLAFAFKYFVQGMGMPSPASMTRLMRSTRASGSSVTSIDGT 340
Db      294 VLERQPAKLAFAFRLFLQGBGYAVGTLOKLARKISSVSRSSSTQIELT 341

```

Search completed: January 29, 2004, 23:29:18
 Job time : 43 secs